Elettronica e telecomunicazioni

Anno LII
Numero 1
Aprile 2003

Editoriale

Analisi della qualità video per applicazioni webcasting

Screensaver sul web

Realizzazione di un DVD video di Rai Teche

Metadati e Modellazione:

Evoluzione nella gestione
dell'informazione nel mondo del broadcaster

Che cosa è, come funziona:

Le origini del video digitale

Utensili per la compressione video

Uno standard pervasivo (MPEG-2 video)

Advanced Video Coding: il prossimo futuro - Parte I

Elettronica e telecomunicazioni

Edizione ottimizzata per la stampa.
La rivista è disponibile su web
alla URL www.crit.rai.it/eletel.htm

Anno LII
N° 1
Aprile 2003

Rivista quadrimestrale a cura della Rai

Direttore responsabile
Gianfranco Barbieri

Indice

Editoriale

ing. Gianfranco **Barbieri**
Direttore di
"Elettronica e
Telecomunicazioni"

Sin dai suoi esordi la radiodiffusione radiofonica e televisiva ha avuto un enorme impatto sugli usi e i costumi della nostra società. L'evoluzione tecnologica che, nel corso di oltre 70 anni, ne ha radicalmente trasformato il livello qualitativo di fruizione (introduzione del colore, radiodiffusione diretta da satellite, TV digitale) non ne ha sostanzialmente modificato il modello di rappresentazione basato sulla trasmissione ad una moltitudine di utenti di una sequenza di immagini con audio associato.

Per la verità, l'introduzione degli standard MPEG ha aperto le porte ad una vasta gamma di possibilità nella erogazione di nuovi servizi radiodiffusi, dalla *near-video-on-demand* alla *pay-per-view*; limitate forme di interattività sono rese fattibili mediante l'utilizzo del canale di ritorno separato.

La recente esplosione commerciale di Internet e la convergenza delle tecnologie stanno, tuttavia, chiaramente indicando che il modello tradizionale potrebbe presto rivelarsi inadeguato per i gusti e le esigenze dell'utente moderno, abituato a servirsi del computer ed a navigare nella rete. Secondo il parere di qualificati sociologi e analisti del mercato, la diffusione, in Italia, di Internet nel prossimo decennio sarà pari a quella della televisione negli anni sessanta e settanta; l'evoluzione del comportamento degli utenti sarà favorita dalla tendenza ad una fruizione meno collettiva (famiglia riunita in salotto attorno al televisore) e più individuale (utente singolo seduto davanti alla consolle con schermo di dimensioni medio-piccole). Questa individualizzazione a livello di utenza sarà favorita dalle tecnologie che permetteranno di fornire servizi sempre più personalizzati e, in tale contesto, è prevedibile una ulteriore evoluzione che vedrà i contenuti multimediali provenire indifferentemente sia dall'etere che on-line, in forma integrata sullo stesso schermo. Tutto ciò richiederà, da un lato, l'incremento della capacità trasmissiva della "Rete", e dall'altro, la codifica di immagini e suono a basso bit-rate e qualità elevata.

L'articolo "Analisi della qualità video per applicazioni webcasting", pubblicato nel presente numero della rivista, tratta una di queste problematiche legate alla misura della qualità nella codifica a basso bit-rate.

L'impegno della Rai nella sperimentazione di nuovi servizi, ed in particolare i servizi on-line, è invece sottolineato nell'articolo che segue ("Screensaver sul web") in cui vengono descritti i criteri seguiti per sviluppare un sito ricco di contenuti grafici e per individuare il miglior compromesso fra la qualità delle informazioni video ed i tempi di accesso e scaricamento.

Oltre che dal "lato utente", l'informatizzazione del sistema radiotelevisivo procede a ritmo crescente anche dal "lato generazione dei contenuti"; in altre parole, l'ambiente della produzione televisiva sta subendo una seconda fase di digitalizzazione, rappresentata dall'evoluzione verso le tecnologie ICT. Esempi significativi provengono, a livello quasi generalizzato, dalle aree server-assistite della produzione, della post-produzione, della messa in onda e dell'archiviazione.

L'articolo "Realizzazione di un DVD video di Rai Teche offre alcune considerazioni sulla versatilità del supporto DVD per l'accesso strutturato alle informazioni audiovisive".

Più recentemente si sta assistendo a tentativi di estendere l'uso delle memorie a disco rigido alle applicazione mobili di *news-gathering* e, più in generale, dell'intera produzione televisiva. Il denominatore comune di tutte queste applicazioni è costituito dal trasporto e memorizzazione dei dati direttamente in formato file. Appare subito evidente che, l'interoperabilità tra apparati provenienti da costruttori diversi e, a maggior ragione, lo scambio di programmi tra radiodiffusori può avvenire senza difficoltà solo se i dati sono formattati secondo protocolli e piattaforme aventi interfacce standardizzate. Di pari passo con l'evoluzione delle metodologie per il trattamento dei dati costituenti il segnale televisivo sta acquisendo un'importanza strategica, nel contesto dell'ottimizzazione delle risorse, lo sviluppo di un sistema di dati associati al programma in fase di realizzazione (i cosiddetti "metadati") contenenti una dettagliata descrizione dei più importanti parametri relativi al contenuto; nella filosofia di impiego del sistema, i "metadati" alimentano un flusso di informazioni tra produttori e distributori di contenuti, gestori degli archivi e utilizzatori finali. La definizione delle caratteristiche funzionali ed i processi di standardizzazione in atto aprono vaste problematiche che vedono impegnati vari organismi internazionali nella elaborazione di linguaggi e protocolli per quanto possibile comuni. Per dare ampio spazio all'informativa sugli sviluppi della materia, la rivista intende con il presente numero iniziare la pubblicazione di una serie di articoli sui temi inerenti la gestione dei dati in una moderna impresa di produzione e diffusione dei contenuti multimediali.

Nella rubrica "Cos'è, come funziona" viene pubblicato un corpo di "tutorials" che costituiscono una panoramica sui principali temi inerenti la codifica del segnale video: codifica del segnale in studio, metodi di compressione dell'informazione, standard MPEG-2 e il nuovo standard AVC - H.264. Su quest'ultimo argomento, particolarmente attuale poiché riguarda un sistema destinato a rivoluzionare ulteriormente il comparto della distribuzione di contenuti multimediali, verranno pubblicati successivi articoli nei prossimi numeri.

Analisi della qualità video per applicazioni webcasting

ing. Paola **Sunna**
Rai
Centro Ricerche e
Innovazione Tecnologica
Torino

Presidente del gruppo
EBU/UER B/VIM

Sommario
Nel 2001, il gruppo EBU B/VIM (Video in Multimedia) è stato creato allo scopo di studiare e definire una metodologia soggettiva per la valutazione della qualità video nell'ambito di applicazioni multimediali. Il contesto si riferisce all'utilizzo di codec per il web (Windows Media, Real Video, Mpeg4.) che, effettuando la compressione del segnale video per consentirne latrasmissione nelle bande tipicamente disponibili su Internet, introducono delle degradazioni sulla qualità fruita dall'utente finale. L'articolo descrive la metodologia proposta dal gruppo EBU, il tipo di test eseguiti ed i risultati preliminari ottenuti.

1. Introduzione

Il termine webcasting è utilizzato per indicare la produzione e trasmissione (streaming, download, video on demand) di contenuti audio, video, ... fruibili dagli utenti finali su terminali (PC, Palmari, PDA…) connessi ad Internet.

Una architettura tipica per il webcasting è riportata in figura 1.

I contenuti audio-video possono essere ripresi dal vivo ("live") oppure essere stati acquisiti e sottoposti ad editing per la creazione di presentazioni "ad hoc" per il web; la codifica è eseguita tramite opportune tecniche di compressione allo scopo di consentire la trasmissione nella banda a disposizione (modem, xDSL, fibra ottica ….) mentre la distribuzione verso gli utenti finali avviene utilizzando appositi server (web server o server di streaming).

Fig. 1 - Architettura Webcasting

Il presente articolo riporta i risultati preliminari ottenuti durante l'esecuzione di una serie di test mirati ad ottenere indicazioni sulla qualità fornita dai principali codec diffusi su web.

Generalmente la valutazione della qualità degli algoritmi di codifica si effettua tramite prove soggettive formali oppure attraverso prove expert viewing; per esempio, la Raccomandazione ITU-R BT.500 [1] riporta le metodologie adottate per la valutazione soggettiva della qualità video in ambito di applicazioni broadcasting mentre, tuttora, non esistono Raccomandazioni che indichino i criteri con cui valutare la qualità dei codec usualmente utilizzati per applicazioni multimediali.

2. B/VIM (Video in Multimedia)

Nel 2001,un nuovo gruppo EBU, il B/VIM [Nota 1], è stato creato allo scopo di studiare e definire una metodologia soggettiva per la valutazione della qualità video nell'ambito di applicazioni di webcasting.

Nota 1 - al progetto B/VIM (Video in Multimedia) dell'UER/EBU hanno aderito oltre alla Rai (il Centro Ricerche e Innovazione Tecnologica), l'IRT (Institut fur Rundfunkentechnik) che è il centro ricerca di televisioni tedesche (ARD,ZDF, DLR), austriache (ORF) e svizzere (SRG/SSR), la televisione norvegese NRK (Norsk Rikskringkasting AS) e FTRD (France Telecom R&D).

I codec introducono dei degradamenti più o meno visibili in funzione del bit-rate di codifica e della complessità del materiale sorgente; i degradamenti possono consistere in perdite di frame con conseguente percezione a scatti del movimento, presenza di blocchettizzazione, perdita di risoluzione e così via. Metodologie oggettive basate sulla valutazione del rapporto segnale rumore non forniscono informazioni relative al degradamento sufficientemente correlate con la qualità percepita dall'osservatore, per cui il metodo più attendibile per ottenere un'indicazione delle prestazioni di nuovi algoritmi di compressione o di nuovi codec resta quello delle prove soggettive.

La necessità di formulare una nuova metodologia, differente rispetto a quelle riportate nella [1], nasce dal fatto che il terminale di fruizione non è più un CRT (Cathode Ray Tube), ma diventa un PC e la risoluzione dei formati non è quella televisiva (Full Format 720x576), ma si riduce (CIF oppure QCIF) per far fronte alla ridotta banda disponibile per le connessioni Internet (Modem, ISDN, ADSL…).

La metodologia soggettiva è stata proposta

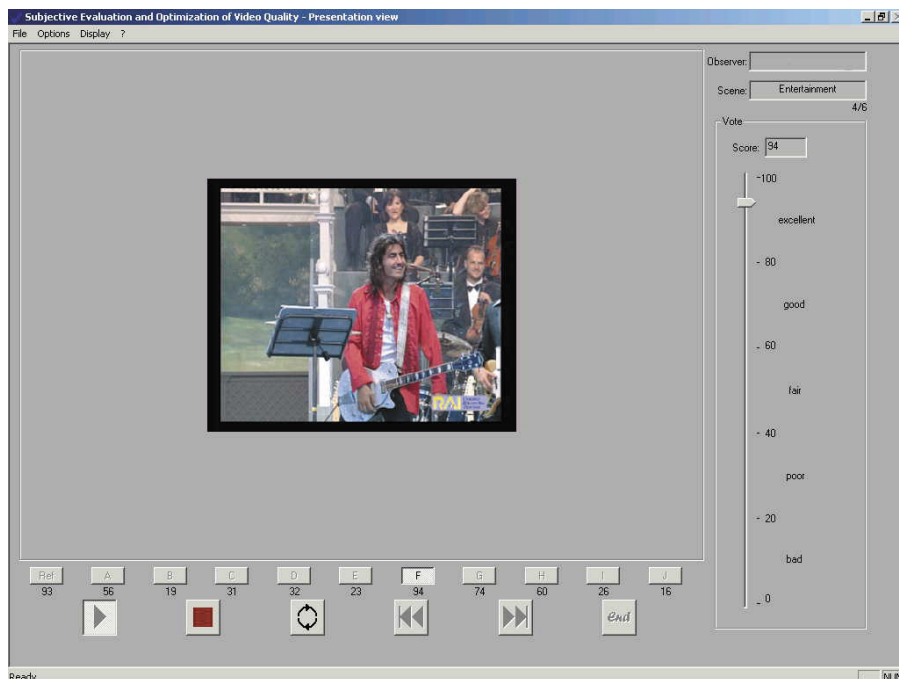



Fig. 2 - Interfaccia grafica per la somministrazione dei test

Connessione	Bit-rate	Bit-rate Effettivo (kbps)	Audio (kbps)	Video (kbps)	QCIF Frame Rate (Hz)	CIF Frame Rate (Hz)
Modem	56	40(+/- 10%)	8 mono	32(+/- 10%)	6.5	
Isdn	128	100(+/- 10%)	20 mono	80(+/- 10%)	12.5	6.5
DSL	256	200(+/- 10%)	32 stereo	168(+/- 10%)	25	12.5
DSL	500	400(+/- 10%)	48 stereo	352(+/- 10%)		25
DSL	700	560(+/- 10%)	64 stereo	500(+/- 10%)		25

Tab. 1 - Parametri utilizzati per le odifiche

da FTRD e si basa su interfaccia grafica su PC (figura 2) che consente all'osservatore di esprimere il suo giudizio di qualità utilizzando una scala continua compresa tra 0 e 100.

Come si può osservare dalla figura 2, l'interfaccia è caratterizzata dalla presenza di una serie di bottoni indicati con Ref ed una successione di lettere A, B, C..

2.1 Descrizione del funzionamento dell'interfaccia

All'osservatore viene chiesto di sedersi e di scegliere la distanza di visione che preferisce. Viene informato che la sequenza associata al tasto Ref (Reference) è quella di riferimento, caratterizzata dalla qualità migliore in quanto rappresenta la sequenza (clip) in formato non compresso. I bottoni (A, B ,C,.) corrispondono alla stessa sequenza codificata con codec e bit-rate differenti.

Tra le sequenze sotto test viene inserito nuovamente il Reference senza che l'osservatore ne sia al corrente allo scopo di verificare la correttezza dell'esecuzione del test, legata al fatto che l'osservatore sia in grado di riconoscere il Reference e di votarlo all'incirca come il Reference esplicito

L'osservatore può rivedere la stessa clip quante volte preferisce e modificare, eventualmente, il suo giudizio.

Fig. 3 - Basket

Fig. 4 - Entertainment

Nota 2 - I metodi di codifica disponibili sulla maggior parte dei codec si distinguono in "single-pass" e "two-pass": quest'ultimo differisce dal primo in quanto il codificatore processa la sequenza due volte: durante il primo passo, il codec acquisisce informazioni sulla complessità del contenuto mentre durante il secondo passo esegue la codifica ottimizzando il processo in base alle informazioni raccolte durante il primo. La codifica "two pass" non può essere utilizzata per codifiche live.

Il tasto >> consente di passare alla sequenza successiva.

Al termine della sessione i dati espressi dai singoli osservatori vengono salvati su un file per essere successivamente elaborati.

2.2 Test condotti dal gruppo B/VIM

Allo scopo di verificare la validità della metodologia proposta da France Telecom, il gruppo ha deciso di selezionare cinque sequenze (Basket, Entertainment, Kayak, News e Flower & Garden) che fossero rappresentative della normale programmazione televisiva, dato che lo scopo non era quello di mettere in crisi i codec sotto test. Le figure 3-4-5-6-7-8 corrispondono ad un fotogramma rappresentativo delle sequenze sotto test.

I codec considerati sono riportati di seguito:

1) Windows Media Encoder 8 (WM8)
2) Real Video 8 (RV8)
3) Sorenson 3
4) Quick Time (MPEG-4)
5) Dicas (MPEG-4)

I dettagli tecnici relativi ai singoli codec sono riportati nelle schede allegate

La tabella 1 indica il bit-rate (audio e video), il frame-rate e le risoluzioni adottate per lo svolgimento dei test.

3. Caratteristiche generali dei codec sotto test

Nell'ambito dei test eseguiti le piattaforme sono state utilizzate per codificare file non compressi in formato AVI; i codec sono stati utilizzati in modalità CBR, un passo [Nota 2].

I codec consentono di specificare una serie di parametri che si ripercuotono sulla qualità finale del filmato e sulla dimensione del file ottenuto.

Tali parametri riguardano:

- audio bit-rate
- video bit-rate
- dimensione del buffer (s): questo parametro deve essere specificato per la codifica CBR in singolo e doppio passo mentre per quella VBR è l'encoder che determina la dimensione ottimale del buffer
- tipo di codifica: CBR, VBR, due passi CBR, due passi VBR
- formato dell'immagine:altezza e larghezza
- frame-rate massimo (Hz)
- distanza tra i key frame (s): rappresenta la distanza temporale tra due fotogrammi chia-

Fig. 5 - Horse Riding

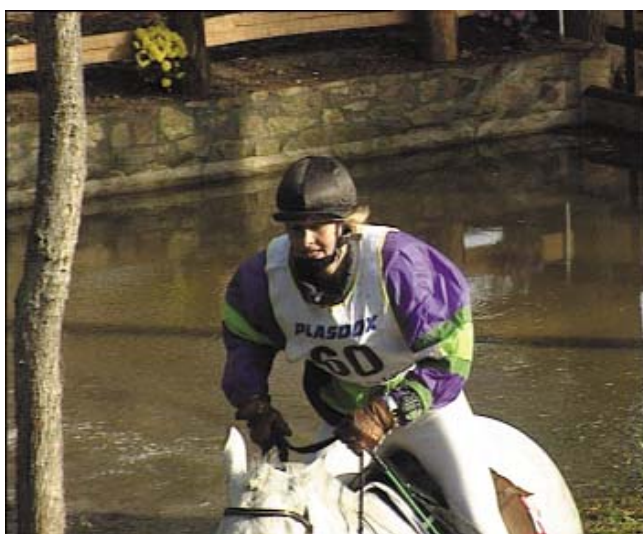



Fig. 6 - Kayak

ve, ossia editabili; i fotogrammi compresi tra due key frame contengono esclusivamente informazioni differenziali rispetto al key frame precedente. Il valore dovrebbe essere pari a pochi secondi per sequenze caratterizzate da elevato movimento e maggiore per sequenze più statiche

- qualità: a seconda del valore assunto da questo parametro, l'encoder privilegia la nitidezza del contenuto o la fluidità del movimento

In generale valgono le seguenti considerazioni:

- il valore associato alla qualità può avere ripercussioni sul frame-rate effettivo di codifica in quanto l'encoder, nel caso in cui i contenuti siano particolarmente critici in termini di dettaglio spaziale e di presenza di movimento, può eliminare alcuni frame (drop) per rispettare le impostazioni relative alla qualità
- il bit-rate (kbps) ottenuto in riproduzione potrebbe non rispecchiare fedelmente quello impostato durante la codifica, ma essere superiore di qualche kbps a seconda delle dimensioni degli header inseriti dal codec
- la dimensione finale del file non coincide con quella ottenibile moltiplicando il bit-rate impostato per la lunghezza della sequenza, ma è superiore di qualche decina di kbyte in funzione della dimensione del buffer e della distanza dei key frame; buffer lunghi e distanze tra key frame brevi provocano un aumento delle dimensioni del file finale rispetto a quelle attese.

4. Analisi della qualità video

L'analisi statistica dei dati forniti dai singoli laboratori, eseguita da FTRD, ha dimostrato la correttezza e la riproducilibità dei risultati ottenuti applicando la metodologia proposta dal gruppo B/VIM.

Di seguito sono riportate:

- le considerazioni derivanti dall'esecuzione di prove expert-viewing, ossia prove soggettive che coinvolgono un numero limitato di osservatori (4-5 persone) esperte nell'ambito della valutazione della qualità video
- i risultati delle prove soggettive formali.

Le expert-viewing sono state eseguite riproducendo localmente i filmati su un PC (900 MHz, 1024x768) e chiedendo agli osservatori di esprimere un giudizio globale, su una scala a cinque livelli (Pessimo, Cattivo, Discreto, Buono ed Eccellente) in funzione della definizione del dettaglio spaziale e della fluidità del movimento.

I risultati emersi dalle expert-viewing sono

Fig. 7 - Teaching

Fig. 8 - Flower & Garden

riportati di seguito:

- WM8 e RV8 forniscono una qualità confrontabile anche se i difetti introdotti dai due codec sono diversi: le sequenze codificate con WM8, soprattutto a bassi bit-rate, presentano una forte blocchettizzazione mentre le sequenze trattate con RV8 sono caratterizzate da una forte perdita di risoluzione nei dettagli
- Sorenson 3 fornisce una qualità nettamente inferiore a quella di WM8 e RV8, soprattutto a bassi bit-rate
- Dicas e QuickTime forniscono qualità confrontabili ed inferiore, rispetto a WM8 e a RV8

Fig. 9 - Risultati delle prove soggettive per il formato CIF

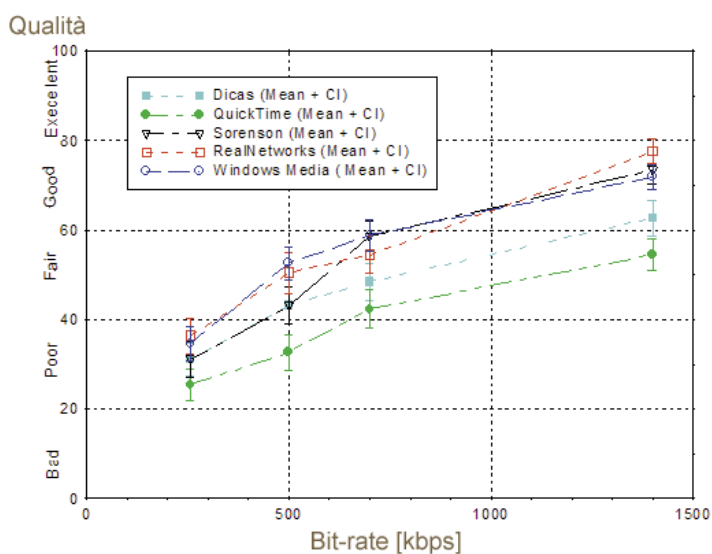



Fig. 10 - Risultati delle prove soggettive per il formato QCIF

Per quanto riguarda la qualità video fornita dagli encoder valgono le seguenti considerazioni:

- per il formato QCIF, almeno 128-256 kbps sono necessari per ottenere un indice di qualità compreso tra discreto e buono
- per il formato CIF , almeno 500-700 kbps occorrono per ottenere un'indicazione di qualità compresa tra discreto e buono

ovviamente tali considerazioni sono strettamente legate al tipo di contenuto che viene codificato per cui a parità di bit-rate la qualità soggettiva percepita può essere fortemente diversa a seconda che il contenuto sia semplice (es. News, Teaching) o più complesso (es. Kayak, Basket).

Le considerazioni scaturite dalle expert-viewing hanno trovato conferma nei risultati delle prove soggettive formali per i formati CIF e QCIF riportati rispettivamente nelle figure 9 e 10.

5. Altri codec per applicazioni Webcasting

Esiste una notevole proliferazione di codec per applicazioni di Webcasting e l'evoluzione del mercato è estremamente rapida. Parecchi produttori offrono soluzioni complete end-to-end che comprendono oltre alla piattaforma di codifica, quella di authoring, streaming e di play-out. La maggior parte delle soluzioni sono "proprietarie" e la tendenza dei costruttori è quella di fornire dei plug-in che consentano di utilizzare Real Video oppure Windows Media come player del materiale da loro codificato.

La tabella 2 riporta un elenco di altri prodotti disponibili sul mercato, alcuni dei quali sono freeware, mentre altri sono a pagamento.

Codec	Società	Sito Web	Architettura
Vp5	On2	www.on2.com	Plug-in per RealVideo
ZyGoVideo	ZyGoVideo	www. ZyGoVideo.com	Plug-in QuickTime
3ivx	Happy Machine	www.3ivx.com	Plug-in per QuickTime
Divx	Divx	www.divx.com	Proprietaria
H.263	Apple	www.apple.com	QuickTime
MPEG-4		www.m4if.org/products	Alcune case produttrici forniscono dei plug-in per Real e/o per QuickTime mentre altre forniscono soluzioni end-to-end proprietarie

Tab. 2 - Codec per il video su web

H.264

Un commento a parte merita l'H.264 (noto anche come MPEG-4/Parte 10), un codec di ultima generazione sviluppato inizialmente dal gruppo ITU-T VCEG (Video Codec Expert Group), successivamente ITU e ISO hanno coordinato i loro sforzi creando il JVT (Joint Video Team) allo scopo di finalizzare uno standard che consentisse di ottenere un guadagno di almeno il 50% rispetto agli standard precedenti. L'approvazione da parte delle due organizzazioni internazionali ed il rilascio della versione definitiva del software è previsto per Maggio 2003. Le prestazioni di questo codec sembrano essere davvero promettenti sia in termini di rapporto qualità/bit-rate che di robustezza su canali in presenza di errori. Informazioni su H.264 sono reperibili in [2] e [3].

La tabella 3 riporta i risultati di alcuni testi condotti dall'HHI su sequenze in formato CIF allo scopo di misurare il guadagno dell'H.264 rispetto ad altri algoritmi di codifica (H.263 Hihg level profile, MPEG-4 Advanced Single Profile e MPEG-2).

E' necessario tenere presente che gli elevati guadagni riportati in tabella 3 sono ottenuti a discapito della notevole complessità computazionale (ved. Scheda allegata per dettagli tecnici sul funzionamento dell'algoritmo) che caratterizza l'algoritmo.

6. Conclusioni

Il mercato dei codec per il web è in continua evoluzione; al momento della redazione dell'articolo sono già disponibili nuovi prodotti software da parte di Microsoft (Windows Media 9), di Real (Real Video 9) e Sorenson (Sorenson Squeeze). Questi nuovi prodotti saranno presi in considerazione dal gruppo B/VIM nella seconda fase del progetto che partirà nel mese di giugno 2003. Alcuni risultati relativi al confronto delle prestazioni tra Windows Media 9 e Real

	Guadagni di bit-rate (Formato CIF)		
Coder	MPEG-4 ASP	H.263	MPEG-2
JVT/H.264	38,62%	48,80%	64,46%
MPEG-4 ASP	-	16,65%	42,95%
H.263 HLP	-	-	30,61%

Tab. 3 - Guadagni dell'H.264. Risultati di test condotti da HHI (Heinrich Hertz Institute) che è fortemente impegnato nello sviluppo del software.

Video 9 sono riportati in [4]: i test sono stati eseguiti da un laboratorio indipendente Keylabs incaricato da Real di condurre la survey.

Per tutti i codec considerati, la qualità video fornita è strettamente dipendente dal contenuto e dai parametri di codifica; tali parametri sono a loro volta legati alla banda a disposizione e consentono ampi margini di libertà.

Un'altra caratteristica comune ai codec nati per applicazioni web è la loro graduale migrazione verso applicazioni di tipo broadcasting (SDTV – Standard Definition Television, HDTV – High Definition Television); Windows Media 9 e Real Video 9, così come l'H.264, si stanno ponendo come possibili candidati per un'eventuale sostituzione di MPEG-2; i guadagni dichiarati dai produttori sono dell'ordine di circa 2 – 3 volte rispetto a MPEG-2 a parità di bit-rate. Al momento della redazione dell'articolo, il Centro ricerche ed Innovazione Tecnologica sta conducendo dei test allo scopo di verificare il guadagno di questi nuovi algoritmi di codifica rispetto a MPEG-2 nel caso di codifiche su materiale SDTV; i risultati saranno oggetto di un prossimo articolo

Bibliografia

1. Recommendation ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures"
2. ITU-T H.26L Standardisation (ITU-T Q6/16, VCEG), www.tnt.uni-hannover.de/plain/project/vceg
3. R. Schaefer, T. Wiegand, H. Schwarz: "The emerging H.264/AVC standard", EBU Technical Review, January 2003, www.ebu.ch/trev_293-schaefer.pdf
4. www.keylabs.com/results/realnetworks/vidperf9.shtml

Acronimi e sigle	
AVI	Audio Video Interleave: formato file
CBR	Constant Bit Rate
CIF	Common Intermediate Format: formato immagine, 352 pixel e 288 righe
EBU/UER	European Broadcasting Union/ Union Européenne de Radio-Télévision (www.ebu.ch)
Frame	Quadro televisivo
Frame rate	Frequenza di quadro: numero di quadri al secondo [Hz]
Key-frame	fotogramma di tipo Intra utilizzato per predire i quadri successivi
ISDN	Integrated Services Digital Network: standard internazionale di comunicazione a 64 kbps
ISO	International Organisation for Standardisation (www.iso.org)
ITU	International Telecomminication Union (www.itu.org) Union Internationale de Télécommunication Unión Internacionale de Telecomunicaciones
MPEG	Motion Picture Expert Group: gruppo di lavoro congiunto ISO/IEC
PDA	Personal Digital Assistant: dispositivo palmare
Q-CIF	Quarter Common Intermediate Format: formato d'immagine, 176 pixel per 144 righe
VBR	Variable Bit Rate
xDSL	tecnologie Digital Subscriber Lines: modem e protocolli, per trasferire dati su doppino

Screensaver sul web

Carlo **Bonugli**,
Andrea **Falletto** e
Mario **Muratori**
Rai
Centro Ricerche e
Innovazione Tecnologica
Torino

Sommario
Screensaver è una trasmissione televisiva che manda in onda video realizzati da ragazzi e da ragazze. L'articolo è relativo alla realizzazione del sito web, basato principalmente sulla tecnologia Flash, a complemento della trasmissione.

1. Introduzione

La programmazione invernale 2002-2003 di RaiTre ha visto la nascita di una nuova trasmissione televisiva mirata agli adolescenti, prodotta presso il Centro di Produzione Rai di Torino (CPTO) e condotta da Federico Taddia per la regia di Paolo Severini.

Lo scopo del progetto è di “salvare la TV”, proponendo una programmazione televisiva scevra dalle distorsioni di cui è affetta la televisione odierna perché prodotta dagli stessi adolescenti, che vengono stimolati a raccontare storie con video prodotti da loro stessi, oppure sono aiutati a realizzare reportage di loro ideazione.

Inoltre, illustrando “come si fa la televisione e di quanto gira attorno ad essa”, si propone un “corso di TV” diluito nelle 180 puntate in cui è articolata la trasmissione.

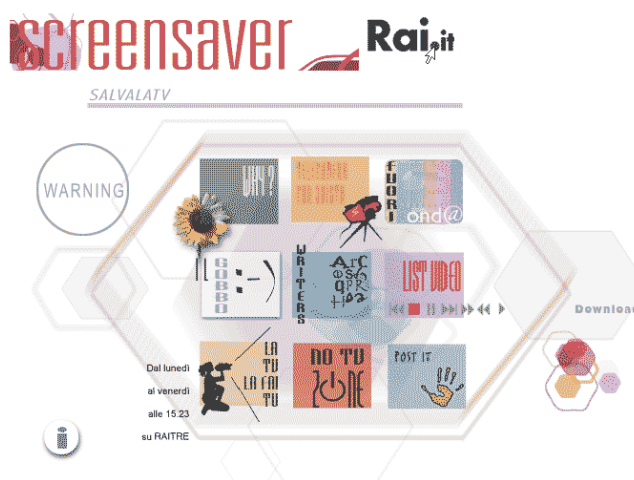
Come ormai consuetudine, alla trasmissione televisiva si è affiancato un sito web, e il Centro Ricerche ed Innovazione Tecnologica (CRIT) è stato coinvolto nel suo sviluppo e nella sua gestione.

2. Le motivazioni

I responsabili del programma hanno ritenuto opportuno associare un sito web alla trasmissione televisiva soprattutto per prolungare il contatto con l’utenza al di fuori degli orari di programmazione.

Già a partire dalla sua ideazione si è tenuto in debita considerazione il fatto che gli adolescenti sono uno dei target più difficili per un sito web, perché alla grande curiosità di “mettere alla prova” il prodotto, generalmente associano un senso di critica elevato e una grande diffidenza verso qualsiasi cosa provenga da soggetti considerati “istituzionali”.

Bisognava quindi fornire un applicativo attraente, in quanto di gradevole aspetto e facile da usare, interessante, perché ricco di contenuti, leggero da scaricare per limitare i tempi di attesa delle pagine, di ampia fruizione sulle più diverse piattaforme; ma anche dotato di meccanismi informatici non banali per non sembrare “inferiore” e, soprattutto, che presentasse il minor numero possibile di malfunzionamenti, nella consapevolezza che sarebbe stato “giudicato” anche dal punto di vista tecnico.

Figure - Questa, e le altre figure a illustrazione dell'articolo, rappresentano le varie sezioni in cui è articlato il sito www.screensaver.rai.it

La sfida si presentò quindi molto impegnativa fin dall'inizio, e conseguentemente molto elevato fu l'interesse del CRIT a partecipare alla realizzazione del sito, anche perché si presentava l'occasione di utilizzare tecnologie più avanzate di quelle adottate in altre precedenti collaborazioni con il CPTO.

3. Progetto e realizzazione

La struttura dell'ipertesto è molto semplice. Prevede undici sezioni, alle quali si accede direttamente dalla "home page", relative ad altrettanti canali di comunicazione con l'utente caratterizzati da contenuti ed informazioni di tipo differente.

Alcune sezioni sono canali a senso unico verso l'utente, sia con contenuto statico, sia permettendo una leggera interattività finalizzata alla fruizione dell'intero pacchetto informativo presente nella sezione.

Altre sezioni propongono invece una interattività più spinta con acquisizione di informazioni dall'utente verso il server, in particolare per la risposta al gioco e per il sondaggio on-line.

In ogni sezione l'interfaccia utente è stata molto curata per renderla gradevole ed interessante, cosa che si è potuta realizzare adottando la tecnologia Macromedia Flash. Quest'ultima infatti, com'è noto, permette di realizzare interfacce grafiche ricche di animazioni e di effetti, molto usato il "rollover" al passaggio del mouse in file di ridotte dimensioni e per questo adatti al loro trasferimento su rete internet anche con connessioni a bassa capacità. Un ulteriore vantaggio della tecnologia Flash è la larga compatibilità coi browser esistenti grazie all'impiego di un apposito riproduttore ("player") gestito come "plugin" dal browser ospite. Inoltre, tra le altre caratteristiche, Flash supporta la programmazione, lo scambio di informazioni col browser e con il server web, che sono state utilizzate anche nel sito di "screensaver" per aumentare la flessibilità di impiego, l'aggiornamento rapido da parte della Redazione del programma e l'interattività spinta con l'utente.

Tralasciando in questo articolo ogni commento sull'ottima realizzazione grafica, gli ultimi tre punti qui sopra elencati richiedono un approfondimento, poiché hanno rivelato non solo delle grandi potenzialità della tecnologia adottata, ma anche dei problemi di tipo tecnico.

La programmazione nell'ambiente Flash è effettuata associando, ad ogni oggetto interessato, uno script nel linguaggio ActionScript, molto simile al linguaggio C. L'ambiente di authoring offre un'interfaccia molto facile da usare, ma dà anche la possibilità di scrivere direttamente lo script per eventuali personalizzazioni. Animare gli oggetti tramite la programmazione permette di alleggerire ulteriormente il file Flash, ma nel caso di "screensaver" non si è ritenuto particolarmente interessante sfruttare questa possibilità, mentre si è utilizzata la programmazione per gestire le comunicazioni col browser ospite e con il web server.

La comunicazione con il browser ospite è stata utilizzata per attivare degli script "lato client" scritti in linguaggio Javascript e gestiti dal browser stesso. Questa è una caratteristica molto potente, ma si è scoperto che non è supportata da alcune versioni di browser su alcune piattaforme, e ciò, per come è stato strutturato il sito, su queste ne impedisce totalmente la navigazione. Per ovviare a questo inconveniente si dovrebbe impostare il progetto del sito in modo da evitare di usare la

comunicazione tra Flash e browser ospite, a meno che i produttori dei browser in oggetto non offrano versioni più complete in cui questa funzionalità sia implementata.

La possibilità di comunicazione tra Flash e il web server è stata utilizzata per modificare il contenuto di alcuni campi che riportano informazioni di frequente aggiornamento. Per meglio comprendere l'importanza operativa di questa potenzialità, bisogna aver presente la procedura di generazione di un file Flash, per lo meno nella versione 5 adottata per il sito. Tale file viene generato come prodotto dell'attività di uno specialista, spesso un grafico, che opera nell'ambiente di authoring di Flash, inserendo nell'applicativo immagini, testi, materiale multimediale, generando animazioni ed eventualmente scrivendo degli script di programmazione. Nel caso in cui si volesse modificare il contenuto di oggetti inseriti nell'applicativo, per esempio il contenuto di campi testuali, si dovrebbe ricorrere al grafico di cui sopra che, operando nell'ambiente di authoring, effettuerebbe le modifiche richieste e produrrebbe un nuovo file da sostituire a quello obsoleto. Il carico del lavoro e il costo conseguente ad ogni modifica risulterebbe quindi molto elevato, se non esistesse una funzionalità che permette all'applicativo Flash di richiedere al web server di scaricargli dei dati. Un programma, attivato sul server in conseguenza a tale richiesta, genera i dati desiderati, nel nostro caso li legge da appositi file, li formatta in maniera opportuna e li trasmette tramite il web server all'applicativo Flash, dove vengono visualizzati. Ciò semplifica enormemente le operazioni di aggiornamento delle informazioni variabili, in quanto non è più necessario ricorrere ogni volta al grafico specializzato, ma è il personale di redazione stesso che prepara il contenuto, nel caso di "screensaver" aggiornato con cadenza giornaliera, e lo memorizza sul server web, per metterlo quindi a disposizione dell'utenza.

Non in tutto il sito però è stato utilizzato Flash. Vi sono infatti alcune pagine, in particolare quella per la partecipazione al gioco e quelle relative al sondaggio, contenenti moduli scritti in HTML che attivano dei programmi residenti sul server web per la gestione dei dati inseriti.

Nel caso del gioco ("Telecamera con svista"), il programma riceve i dati inseriti dall'utente, aggiunge i dati relativi alla data e ora di elaborazione e memorizza il tutto su un apposito file protetto. L'ora di elaborazione è un dato importante in quanto, secondo regolamento, vince il gioco la persona che ha dato la prima risposta esatta ricevuta. Il fatto che tutti i contributi degli utenti debbano obbligatoriamente essere elaborati da un unico programma, attivato su un unico server, dà la garanzia di ottenere un ordinamento temporale corretto su cui basare l'individuazione del vincitore. La Redazione, in questo caso, non fa altro che analizzare i dati memorizzati e procedere secondo il protocollo previsto dal Regolamento.

Il sondaggio ("No TV zone") prevede l'attivazione di due programmi "lato server". Il primo è utilizzato per generare la pagina che mostra i risultati del sondaggio tramite un istogramma a barre. Questa è costruita in modo da essere contemporaneamente anche un modulo atto a ricevere la preferenza dell'utente; di conseguenza alla votazione, si attiva un secondo programma sul server web che aggiorna i dati del sondaggio memorizzati in un apposito file. In questo caso la gestione del sondaggio vero e proprio è completamente automatica, e la Redazione ha il compito di analizzare le proposte alternative fornite dagli utenti (campo: "altro") ed eventualmente riavviare il sondaggio proponendo nuovi elementi di votazione.

Essendo il sito legato ad una trasmissione televisiva, non poteva mancare il materiale video. In questo caso si è scelto di mettere a disposizione una selezione dei filmati prodotti dai ragazzi e proposti nel corso della trasmissione.

A causa dell'elevato numero di filmati (circa 160), la selezione del materiale da visualizzare avviene in tre passi. Nel primo si seleziona un mese e viene visualizzata la lista dei filmati messi in onda nel mese selezionato; nel secondo passo, la selezione di un filmato ne visualizza la scheda relativa; nel terzo si comanda lo scaricamento, o la visualizzazione in modalità "streaming", tramite una serie di "bottoni" nei quali viene anche indicata la dimensione del file e la durata di scaricamento con un modem a 28.8 kb/s.

Si impone a questo punto una considerazione relativa al materiale messo a disposizione sul sito.

Le opportunità di scelta tra differenti codificatori e contenitori offerte dall'attuale sviluppo tecnologico sono molto ampie, differenziandosi per caratteristiche tecniche e accoglienza presso l'utenza.

Data l'impostazione del sito – e della trasmissione - verso una popolazione giovane, innovativa ed esigente, si è ritenuto opportuno offrire materiale di elevata qualità, proponendo una scelta tra un certo numero di codificatori, ognuno con una sua peculiarità che dovrebbe essere apprezzata.

Si è quindi utilizzato il codificatore DiVX, legato ad un mondo informaticamente "alternativo", l'H.264, standard pubblico talmente nuovo che sarà ufficializzato solo nel corso della prossima estate, nonché un codificatore "tradizionale" – per quanto riguarda il video sul web – quale il RealVideo, anche se nella sua forma più avanzata (RealOne con tecnologia Helix).[Nota 1]

Un buon prodotto si è ottenuto riducendo la dimensione dell'immagine a 352x256 pixel, la frequenza di quadro a 15 Hz, e impostando un bitrate del segnale codificato a 450 kb/s.

Ai filmati ad elevata qualità codificati con questi parametri, si sono affiancati filmati adatti alla visualizzazione in “streaming” codificati ad una qualità soggettiva inferiore, adatta per connessioni a partire da 12 fino a 128 kb/s, utilizzando tecnologia RealOne Helix in modalità “SureStream”. Quest’ultima permette alla coppia client (sulla macchina utente)/server (presso Rai) di comunicare tra loro adattando istantaneamente il flusso di dati alla capacità del canale realmente a disposizione.

In tal modo si possono servire anche gli utenti che dispongono di un collegamento verso internet di limitata capacità. Infatti, i filmati codificati ad elevata qualità in genere richiedono tempi di scaricamento su linea telefonica normale non compatibili con le bollette che un “buon padre di famiglia” è disposto a pagare per questo tipo di cose. Sicché la modalità di fruizione in “streaming” è sembrata quanto meno opportuna, nella considerazione che la maggior parte degli utenti abbia a disposizione una comune linea telefonica, essendo molto probabile che i canali ad alta velocità ISDN siano appannaggio di uffici e attività produttive, e che l’ADSL sia ancora pochissimo diffuso, soprattutto nelle aree extraurbane.

4. Un’esperienza in evoluzione

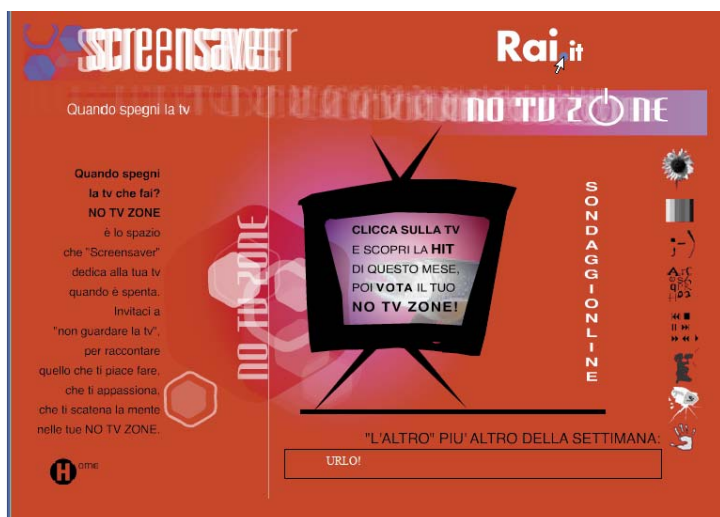
La scelta di basare l’ipertesto su pagine HTML contenenti l’applicativo Flash ha permesso una grande flessibilità e l’armonizzazione grafica tra le pagine delle sezioni, costruite attorno all’applicativo Flash corrispondente, e le pagine in HTML usate per funzionalità secondarie (gioco e sondaggio). Infatti tutte le pagine riportano in testa un applicativo Flash contenente il titolo del programma, un altro per il banner pubblicitario, e un elemento grafico per il link al sito della Rai; nel corpo della pagina, a seconda del caso, trovano posto l’applicativo Flash “principale” o una pagina scritta completamente in HTML.

L’uso di Flash per le parti “complicate” dell’interfaccia e di “semplice” HTML per la restante parte, ha inoltre permesso una elevata compatibilità con piattaforme e browser differenti. Tuttavia, come sopra accennato, questa soluzione necessita della comunicazione tra applicativo Flash e browser che non è implementata in certe versioni di browser su certe piattaforme.

Per ovviare a questo problema, e aumentare la compatibilità e quindi l’utenza potenziale, la soluzione potrebbe essere di sviluppare tutto l’ipertesto completamente in Flash.

Un problema non ancora risolto è costituito dalla qualità del materiale video che è possibile, con la tecnologia attuale, trasferire all’utente in tempi e modi accettabili. Le capacità dei canali a disposizione degli utenti sono limitate rispetto a quanto richiesto dal segnale televisivo digitale in standard ITU-R BT.601, che, quindi, deve essere elaborato per ridurne la richiesta in termini di capacità di canale di trasmissione, se trasferito in modalità “streaming”, oppure in termini di tempo di trasferimento se scaricato come file. In genere, le tecniche utilizzabili prevedono di ridurre le dimensioni dell’immagine e la frequenza di quadro, ed effettuare una compressione “con perdita”. La scelta dei parametri più opportuni è frutto di un compromesso tra dimensione del file e qualità soggettiva, tenuto conto dell’uso previsto del materiale.

Nota 1 - Si noti che solamente lo H264 è uno standard “pubblico”, essendo stato sviluppato nell’ambito dell’ITU (International Telecommunication Union), anche se il codificatore utilizzato è stato prodotto da un’azienda, mentre sia il DiVX, sia il RealVideo sono attualmente standard proprietari, ancorché molto diffusi su internet (anche perché i visualizzatori, “player”, sono di norma disponibili in forma gratuita).

Per il sito di screensaver si è voluto dare risalto alla qualità dei filmati evitando di penalizzarli con una codifica a qualità soggettiva eccessivamente ridotta. Ciò ha portato, però, a file di dimensioni notevoli, che possono essere agevolmente scaricati solo da utenti che hanno a disposizione canali ad elevata capacità quali l'ADSL. Per servire anche gli altri utenti si è ricorso alla codifica di tipo "streaming", modalità "multilivello", che offre una qualità dipendente dal canale trasmissivo a disposizione, ma permette la fruizione del materiale anche agli utenti connessi in rete con le normali linee telefoniche.

Un altro problema emerso già nei primi giorni è legato alla posta elettronica. Questa, infatti, è un mezzo ormai comunemente utilizzato, soprattutto con chi ha dimestichezza con internet, ed in particolare il target di riferimento del programma, quindi quanto di meglio per instaurare un canale di interazione programma-utente. Per questo motivo, nell'interfaccia grafica si sono inseriti dei link speciali che attivano il programma di posta elettronica residente sulla macchina dell'utente, al fine di facilitarne l'uso. La quantità di lamentele sollevate dagli utenti ha evidenziato come moltissimi utilizzino non già la funzionalità di posta di internet gestita con apposito client, bensì i servizi di posta elettronica messi a disposizione da molti internet provider, e accessibili via browser, annullando perciò l'utilità degli automatismi di cui sopra e spingendo a ricercare differenti forme di interattività.

5. Riconoscimenti

Il sito web di "screensaver" è frutto della collaborazione tra il Centro di Produzione Rai di Torino e il Centro Ricerche ed Innovazione Tecnologica della Rai, col supporto tecnico del servizio ICT.

La realizzazione e l'aggiornamento del sito sono stati possibili grazie all'intensa e fattiva collaborazione tra la responsabile di produzione Mussi Bollini, gli autori Federico Taddia,

Massimo Bruno ed Elena Mora, la Redazione, con particolare riferimento a Cristina Cuzzupoli, la grafica Simona Castagnotti, ed il CRIT.

Presso il CRIT la programmazione (Flash, Javascript, PHP) e l'integrazione dei vari contributi sono state curate da Mario Muratori, mentre le codifiche video sono state effettuate da Carlo Bonugli e Andrea Falletto.

Realizzazione di un DVD video di Rai Teche: considerazioni sulla versatilità del mezzo

Marzio **Barbero** e
Andrea **Falletto**
Rai
Centro Ricerche e
Innovazione Tecnologica
Roberto **Rossetto**
Rai
Teche

Sommario
E' stato realizzato dalla Direzione Teche il documentario "viaggio nella memoria televisiva della Rai". L'articolo illustra la realizzazione di un DVD basato sul documentario ed in cui vengono sfruttare le caratteristiche di versatilità del mezzo per consentire una semplice fruizione dell'elevato numero di "capitoli" in cui è organizzato il DVD, corrispondenti ad una selezione di video che ripercorrono i momenti più significativi ed emozionanti della storia della televisione.

1. Introduzione

Questo breve articolo ha lo scopo di illustrare con un esempio pratico, le caratteristiche di versatilità del supporto DVD[Nota 1] per l'accesso, in modo strutturato ed efficace, alle informazioni audiovisive.

2. I due campioni a confronto

Nel dicembre 2002 la Direzione Teche della Rai ha chiesto la collaborazione del Centro Ricerche e Innovazione Tecnologica per la realizzazione dell'editing e dell'authoring necessario per produrre un DVD video destinato alla diffusione del documentario "*viaggio nella memoria televisiva della Rai*" di Giancarlo Governi e Barbara Scaramucci.

E' stata richiesta la collaborazione del Centro Ricerche poiché i tempi erano stretti ed era richiesta particolare cura nella realizzazione, per garantire una buona qualità video ed una interfaccia utente semplice, ma efficace: già in passato collaborazioni finalizzate alla realizzazione di prodotti simili si erano dimostrate proficue.

La storia di Rai Teche, la sua missione, il percorso che ha portato alla realizzazione del catalogo multimediale, la valorizzazione dei materiali d'archivio. Questo è anche il criterio ispiratore del documentario[Nota 2] . Queste frasi descrivono lo scopo del documentario e nel testo che lo accompagna e lo integra è ricordata la collaborazione tecnologica[Nota 3] del Centro Ricerche e della direzione ICT nella realizzazione del catalogo multimediale.

Il documentario è stato distribuito in una confezione costituito da una cassetta VHS e dal DVD-video: questa è una opportunità per mettere a confronto i due media. Il formato DVD negli ultimi tempi ha conquistato quote sempre maggiori nel mercato della distribuzione dei prodotti audiovisivi, a spese del formato VHS. Le ragioni sono legate alle forti riduzioni di prezzo dei lettori DVD, mentre la qualità video e audio e la versatilità del mezzo giustificano ampiamente la differenza di prezzo con cui i due prodotti vengono commercializzati.

Nota 1 - DVD è acronimo di Digital Versatile Disc. La scheda "Che cosa è, come funziona: il disco ottico versatile (DVD)" pubblicata nel precedente numero di Elettronica e Telecomunicazioni fornisce alcune delle motivazioni di questa versatilità (il supporto è stato ideato per una gamma di applicazioi, oltre al DVD-video, vi sono i formati DVD-ROM e DVD-audio) e alcuni dati sulla penetrazione di questo medium nel mercato mondiale.

Nota 2 - In *corsivo* è riportato parte del testo che è stato utilizzato dagli autori del documentario per commentare il prodotto da essi ideato.

Nota 3 - Il catalogo multimediale è stato oggetto di due articoli pubblicati su Elettronica e Telecomunicazioni dell'aprile 2000.

3. Esempio di versatilità

Un viaggio in cui, attraverso una ragionata selezione di immagini si ripercorrono i momenti più significativi e le grandi emozioni della storia della televisione, dagli albori fino ai nostri giorni. Questa frase di presentazione del documentario ha una implicazione tecnica significativa: il disco è organizzato in 99 capitoli corrispondenti ai clip video attraverso cui si svolge il viaggio.

Il DVD consente di fruire delle immagini in flusso continuo, così come è possibile utilizzando la cassetta VHS, ma soprattutto permette di accedere ai singoli clip video mediante un insieme articolato di menu e sottomenu, alcuni dei quali rappresentati nelle figure.

4. Peculiarità realizzative

Dal punto di vista tecnico e realizzativo si possono elencare queste considerazioni:

- Nel processo produttivo, la realizzazione del DVD spesso è la fase finale e quindi risente dei ritardi che si sono accumulati lungo la catena: i tempi per attuare le diverse fasi della progettazione sono stretti e i margini per recuperare eventuali imprevisti o errori tendono ad annullarsi.
- Progettazione dell'interfaccia e della navigazione: immediata e di facile uso, i singoli clip sono rappresentati sia da un titolo che da una immagine significativa del contenuto.
- Mentre il committente richiede normalmente una codifica al massimo bit-rate, sinonimo di massima qualità video, il progettista in questo caso ha dovuto considerare lo spazio occupato da menu, dati per la navigazione, audio e video. In questo caso il video ha una durata totale di 78 minuti, organizzati in 99 capitoli e 16 menu e sottomenu. Il video è stato codificato MPEG-2 in modalità CBR (Constant Bit Rate) a 6 Mbps. La codifica è stata effettuata con un sistema hardware di tipo professionale e questo

Figure - Questa figura è relativa al menu principale, mentre le successive si riferiscono al alcuni delle 15 pagine di sottomenu in cui è articolato il DVD.

ha consentito di ottenere una qualità trasparente. L'audio è stato codificato in PCM lineare a 16 bit e con frequenza di campionameno pari a 48 kHz.

- Lo standard DVD è rigido per facilitare la compatibilità dei lettori, ma spesso ciò limita la creatività del progettista. In questo caso è stato necessario limitare il numero dei capitoli a 99 (numero massimo previsto dallo standard), concordando le opportune modifiche al contenuto.
- Limitazioni dei sistemi di authoring: durante la progettazione di un DVD è necessario considerare le ulteriori limitazioni del software di authoring (in genere il costo del software cresce esponenzialmente con l'aumentare delle possibilità di utilizzare appieno lo standard).

5. Conclusioni

Un percorso che, anche graficamente, illustra la "rivoluzione digitale" in corso e vuole dimostrare che, come scrisse Carlo Levi, "il futuro ha un cuore antico". Questa frase conclude il testo introduttivo al documentario.

Analogamente la fortuna del mezzo di distribuzione DVD, destinato probabilmente nel prossimo futuro a subire miglioramenti significativi grazie all'aumento della capacità dati, di cui si è scritto nel numero precedente di Elettronica e Telecomunicazioni, e al futuro standard di codifica video, di cui si parla in questo numero, trae origine dallo standard ITU-R BT.601, prodotto nello scorso millennio, l'ormai lontano 1982.

Metadati e Modellazione
Evoluzione della gestione dell'informazione nel mondo dei broadcaster

ing.. Laurent **Boch** e
ing. Alberto **Messina**
Rai
Centro Ricerche e
Innovazione Tecnologica
Torino

1. Introduzione e motivazioni

1.1 Scopo dell'articolo

Questo breve articolo si pone come introduttivo alla serie proposta, e presenta un'analisi di scenario della corrente evoluzione alla quale si assiste nel campo della gestione dell'informazione all'interno di una moderna impresa di produzione e diffusione di contenuti audiovisivi. Questa introduzione fungerà da contesto per trattazioni successive riguardanti i dettagli dei singoli aspetti qui presentati.

1.2 Metadati: alla ricerca del significato

E' ormai all'ordine del giorno l'utilizzo del termine "metadati" Nota [1] come punto di partenza assodato e come parola chiave convenzionale nei più disparati ambiti. In particolare poi, sempre più frequentemente diviene oggetto di analisi di problematiche correlate alla gestione e sfruttamento dei contenuti audiovisivi.

Nota 1 - La parola *metadato* ha etimologia mista dal greco "meta" (oltre, al di là) e dal latino "datum".

Questo termine insieme a tutta la rete di concetti correlati, si presenta in realtà con sfaccettature semantiche molto numerose.

Noto e usato già nell'ambito bibliotecario espande il suo raggio di valenza semantica alle risorse web fino ad arrivare con elevato impatto nel mondo dei moderni broadcaster e a mille altri ambiti ancora. Ecco alcune delle definizioni reperibili in seguito a semplici ricerche sul web per il termine inglese "metadata" che evidenziano il grado di polisemia che il termine può assumere nei diversi ambiti:

- "[I metadati sono] *dati a proposito dei dati, per esempio possono veicolare informazione a proposito della formattazione dei dati*"
- "[I metadati sono] i*nformazioni a proposito dei dati in sé: tipicamente informazioni a proposito dei dati audio e video presenti in uno flusso binario*"
- "[Un metadato è] *una risorsa che porta informazione a proposito di un'altra risorsa*"
- "[Un metadato è] *un'informazione a proposito di un dato, più specificatamente un'informazione descrittiva fornita attraverso una codifica a tag all'interno di un documento HTML o XML*"
- *"Contenuto = Essence + Metadati"*
- *"[…] Metadato è la definizione o descrizione di un dato."*
- "[Un metadato] *è un dato ausiliario che fornisce informazioni per una interpretazione intelligente di un insieme elementare di dati"*

In generale si potrebbe quindi dire che i metadati sono informazioni riguardanti qualcos'altro, che di volta in volta è l'oggetto principale del nostro interesse all'interno di un contesto ben definito.

Per questi motivi non si vuole dare qui una definizione ultimativa del termine “metadati” poiché si crede che la forte dipendenza dal contesto porterebbe a perdere di vista quello che si ritiene essere il vero nocciolo della questione: l’informazione e il suo significato. Si sposterà quindi il fuoco dell’attenzione dal semplice termine verso ciò che a livello sostanziale ne costituisce l’origine e la ragion d’essere, cioè *l’informazione* e la sua formalizzazione in modelli (cioè entità, relazioni e procedure dettate da regole di business ben identificate) relativi alla conoscenza dei processi e degli oggetti propri di un dominio. I metadati assumono quindi la veste di forma codificata dell’informazione tipica di un dominio, completando ciò cui si riferiscono, arricchendolo e spesso permettendone una gestione efficiente, e diventando a volte elementi indispensabili per la corretta fruizione. Ma occorre tener presente che i metadati in sé sarebbero privi di significato se non si considerasse l’intreccio di relazioni che li lega tra loro e alle entità cui si riferiscono e il contesto stesso in cui i metadati sono definiti. Si restringerà quindi l’attenzione ad un ben particolare dominio, quello tipico di un broadcaster/produttore di contenuti audiovisivi.

Nel seguito sarà presentata un’analisi preliminare a proposito dell’evoluzione che il trattamento dell’informazione sta subendo negli ultimi anni in questo ambito, gettando le basi per future trattazioni più esaurienti e dettagliate dedicate ai singoli aspetti.

2. Evoluzione del modo di utilizzare l’informazione

2.1 Il paradigma tradizionale

Nel paradigma tradizionale di funzionamento di una tipica organizzazione di diffusione e produzione di contenuti, lo spazio di esistenza dell’informazione sotto forma di dati si concretizza in contesti limitati, come ad esempio la documentazione del materiale d’archivio e il reperimento del materiale dall’archivio stesso sulla base della documentazione. Lo scambio informativo tra le varie funzioni (archivio, produzione, ideazione ecc.) è scarsamente automatizzato. Sono possibili livelli di integrazione locali alle singole funzioni (dizionari dati condivisi, metodologie di gestione consolidate, applicazioni integrate) ma scarso rimane il livello d’integrazione con la produzione, l’area acquisti/ideazione, la commercializzazione.

La motivazione di ciò è da ricercarsi preminentemente nella scarsa necessità di questa integrazione, dati i modelli di business dominanti, per un’organizzazione che operi secondo questo paradigma.

I metadati in questo caso codificano essenzialmente l’informazione relativa alla descrizione e all’identificazione degli oggetti e dei concetti che hanno rilevanza nel contesto dell’archivio.

Dal punto di vista della realizzazione dei sistemi, inoltre, può essere ravvisato un altro punto critico in questo approccio tradizionale. Questo prevede, nella maggioranza dei casi di sistemi esistenti, la progettazione e realizzazione di sistemi monolitici all’interno dei quali sono concentrate le informazioni sotto forma di dati.

Data la intrinseca frammentazione di competenze che si realizza in qualsiasi organizzazione riguardo alla gestione delle informazioni, il risultato di questa filosofia si rivela duplicemente carente. Da una parte la gestione dei sistemi è concentrata spesso in un’unica struttura organizzativa, con conseguente inefficienza in termini di comunicazione tra chi gestisce e chi controlla effettivamente l’informazione. Dall’altra parte si tratta spesso di informazioni naturalmente estranee al contesto di chi le deve detenere e gestire, per cui si finisce per costruire basi di conoscenza parzialmente carenti o inconsistenti, con conseguente incompletezza o duplicazione dell’informazione.

Tuttavia, su questo scenario sono pronti ad agire alcuni motori evolutivi ad elevatissimo impatto che riguardano la vita delle imprese che si dedicano alla produzione e diffusione di contenuti audiovisivi; impatto che riguarderà in

maniera sostanziale il modo di concepire ed usare i "metadati".

2.2 I motori dell'evoluzione

I principali motori che in maniera concomitante portano all'evoluzione del paradigma tradizionale sono da individuarsi in:

- Necessità di un abbassamento dei costi di produzione dettata principalmente da esigenze di espansione del mercato e di evoluzione della libera concorrenza.
- Salvaguardia e preservazione dell'archivio. L'evoluzione culturale degli ultimi anni porta a considerare gli archivi in generale e quelli audiovisivi in particolare come elementi fondamentali per la preservazione dell'identità culturale e della storia di una comunità o nazione.
- Abbattimento dei costi di accesso all'archivio. Questa esigenza è in gran parte derivata da quella di abbattimento dei costi di produzione, poiché gli archivi, soprattutto se dotati di elevata estensione temporale, costituiscono una sorgente preziosa di contenuti per la produzione di nuovi programmi.
- Obsolescenza delle tecnologie e ricambio del personale. Spesso l'evoluzione tecnologica porta con sé nuove funzionalità per gli strumenti adottati all'interno delle varie funzioni, ma questo implica che il *know how* debba essere aggiornato continuamente e con velocità sempre più sostenuta. Il ricambio generazionale del personale e la maggior dinamicità del mondo del lavoro rende questo scenario molto critico.
- Esigenza di conquistare e mantenere nuovi mercati (Internet). L'interattività e la multimedialità del mondo Internet apre orizzonti nuovi per lo sfruttamento dei contenuti audiovisivi.

2.3 Le risposte

Una delle risposte naturali alle esigenze di abbassamento dei costi di produzione è ovviamente la riduzione dei tempi di produzione, o in altre parole l'aumento dell'efficienza. Questo è considerato un aspetto di primaria importanza in quanto il tempo impiegato dal personale di produzione è una delle voci che più incide sul costo totale. L'abbassamento dei tempi è perseguibile primariamente attraverso l'innovazione tecnologica, e questo è possibile in alcuni modi:

- Innovazione tecnologica nei processi: normalmente l'avvento delle nuove tecnologie riduce i tempi delle sottoattività critiche[Nota 2], grazie all'introduzione di livelli di automazione nei processi. Le nuove tecnologie riguardano sia gli strumenti di editing e post-produzione, sia gli apparati di ripresa sia i formati digitali di archiviazione e di scambio del materiale. Questo ultimo aspetto ha grande importanza per quanto riguarda la preservazione dell'archivio. In particolare attraverso:
  - o La definizione e l'introduzione di nuovi algoritmi di codifica dell'audiovisivo, per aumentare l'efficienza di immagazzinamento del materiale e quindi diminuire i costi di preservazione dell'archivio.
  - o La definizione e l'introduzione di formati file standard di tipo evoluto, per aumentare l'efficienza degli scambi di materiale durante le fasi del processo di produzione. L'uso di formati informatici abbassa, in generale, i costi di produzione per i nuovi canali di pubblicazione (per esempio Internet, video on demand), rendendo sostenibile la colonizzazione dei nuovi mercati.
- Innovazione tecnologica nell'informazione: parallelamente ai processi l'innovazione tecnologica investe anche l'area dell'informazione, e in particolare aspetti che possono essere classificati sotto la categoria "ingegnerizzazione dell'informazione":
  - o Nuovi modelli informativi: la finalità in

Nota 2 - Un esempio: si pensi al radicale cambiamento di modus operandi che è stato introdotto dal cosiddetto "editing non lineare" o "editing ad accesso casuale" sugli audiovisivi, cioè dalla possibilità di inserire, cancellare e modificare parti senza dover adoperare un procedimento sequenziale. Questa modalità è possibile grazie all'adozione di stazioni di lavoro interamente informatizzate.

questo caso è l'abbattimento dei costi di accesso "qualificato" all'archivio; una rappresentazione più accurata e completa e più rapidamente disponibile dell'informazione descrittiva e identificativa degli oggetti e dei concetti tipici del dominio, permette, ad esempio, il reperimento del materiale d'interesse dall'archivio minimizzando i mancati ritrovamenti o il rumore di ricerca. Inoltre l'applicazione di metodi di ragionamento automatico o metodi dell'intelligenza artificiale alla documentazione e alla classificazione automatiche, permette di effettuare ricerche complesse in tempi ridotti e con costi di popolamento delle basi dati contenuti.

- Nuove tecnologie per lo scambio e la persistenza dell'informazione: la finalità è aumentare l'efficienza dei processi di scambio di materiale tra le varie funzioni e abbattere i costi di sviluppo delle applicazioni, il cui ciclo di vita tende ad essere sempre più breve in risposta all'enorme dinamismo presente nel mondo dell'Information Technology. L'utilizzo di protocolli standard per l'accesso alle basi dati e formati standard per lo scambio di documenti minimizza l'impatto relativo alla gestione della dinamica dei requisiti utente.

- Standardizzazione delle architetture applicative: l'adozione di schemi applicativi standard[Nota 3] riduce i tempi e i costi di progettazione e sviluppo delle nuove applicazioni.

- Evoluzione tecnologica dei sistemi informatici: l'utilizzo di architetture hardware standard reperibili sul mercato a basso costo (come ad esempio personal computers e infrastrutture per reti locali IP) contribuisce ad abbassare i costi di gestione della produzione e rende possibile a livello economico la preservazione di ingenti archivi.

- Conoscenza registrata dei processi (workflow management): la crescente flessibilità del mondo del lavoro e il fenomeno del ricambio del personale rende necessario che, per sostenere i flussi di produzione imposti dai nuovi mercati, i processi tipici delle varie funzioni siano formalizzati e controllati il più possibile in maniera automatica.

Per adattarsi ai nuovi impulsi, quindi, l'impresa è sottoposta a questa vera e propria rivoluzione nel proprio modo di operare. Alcuni interessanti aspetti riguardanti la dinamica di questa risposta saranno analizzati nei seguenti paragrafi.

3. I riflessi dell'evoluzione

3.1 Evoluzione verticale e orizzontale dell'impresa

L'attivazione delle varie risposte ai motori dell'evoluzione implica una profonda trasformazione a livello organizzativo per l'impresa che tratti di produzione e pubblicazione dei contenuti audiovisivi. Quello a cui in generale si assiste è una integrazione a livello impresa della conoscenza aziendale dei processi e dell'informazione.

Si hanno due direzioni ortogonali per questa:

- **Evoluzione verticale**

  Coinvolge le singole funzioni (archivio, produzione, ecc) oppure le singole unità organizzative. Si procede all'ottimizzazione dei sottoprocessi, eventualmente trascurando i vincoli derivanti dal coinvolgimento di altre unità, e si identificano chiaramente le soluzioni di propria pertinenza che sono vicine allo stato dell'arte ed in grado di fornire le prestazioni desiderate.

- **Evoluzione orizzontale**

  Riguarda l'integrazione interfunzionale dell'intera organizzazione, al fine di ottimizzare il business complessivo. Deve quindi risolvere, possibilmente in modo

Nota 3 - Per esempio l'architettura J2EE (Java 2 Platform, Enterprise Edition). E' una piattaforma basata sul linguaggio di programmazione Java progettata per l'elaborazione dati a livello di grande impresa. Sun Microsystems e altri partners tra cui IBM hanno definito questa piattaforma per semplificare lo sviluppo delle applicazioni attraverso l'uso di componenti modulari e standardizzati.

armonico, i problemi di definizione d'interfacce, anche promuovendo la condivisione di strumenti ed informazioni.

La piena attuazione dei due tipi di evoluzione porta allo scenario definibile come integrazione a livello ***"Enterprise"*** (E) al pieno grado di maturità.

L'organizzazione della gestione dell'informazione nel caso ***"Enterprise"*** dovrebbe permettere di specificare in ogni punto, sia quale utilizzo è previsto, sia quali sono gli utilizzatori.

Questo risultato può essere raggiunto comunque in diversi modi, sia attraverso una serie di prese di coscienza ed aggiustamenti successivi, sia perseguendo gli obiettivi di una consapevole pianificazione strategica. Ne consegue che le possibilità di controllo, in vista di adeguarsi alle continue novità, possono essere sensibilmente diverse.

3.2 Approcci alla gestione delle informazioni nel campo audiovisivo

Nel caso dell'industria televisiva le esigenze d'innovazione tecnologica e modernizzazione generale hanno trovato due fondamentali campi d'applicazione.

Uno è quello comune alle imprese di ogni genere e riguarda l'esteso utilizzo delle tecnologie informatiche. Lavorare prevalentemente al computer è diventato la normalità per molte figure professionali, questo però non elimina la complessità di gestire in modo ordinato e coordinato l'intera attività di una azienda, specialmente quando dall'organizzazione adottata dipende l'efficienza complessiva.

L'altro riguarda l'introduzione delle tecniche numeriche nelle generazione e lavorazione dei segnali audio e video, che sono il materiale essenziale del prodotto televisivo. Si tratta in questo caso di una graduale sostituzione dell'intera infrastruttura analogica, che pur è stata frutto in passato di notevoli investimenti.

Ultimamente, però, i nuovi sistemi per un numero crescente di applicazioni audio video derivano più dalla tecnologia informatica generale che non dalla tecnologia puramente televisiva tradizionale (anche se convertita al numerico). Questi sistemi sono caratterizzati dal combinare insieme il fatto di affacciarsi alle reti dati generiche, con tutto ciò che ne consegue in termini di scambi d'informazione e di uso di applicazioni standard, e il fatto di potersi interfacciare con gli apparati solo audio-video. Si notano inoltre una crescente importanza del software rispetto all'hardware e la gestione e lo scambio del materiale video in forma di file.

Per la produzione televisiva si sta passando inesorabilmente dall'avere sistemi numerici/informatici autonomi, all'interno di una catena produttiva tradizionale, alla definizione di una catena produttiva totalmente numerica ed integrata in un flusso di processo più snello e semplificato.

In questo contesto, quale destino si profila per il modo di creare, gestire e utilizzare le varie informazioni associate ai prodotti e al materiale audiovisivo ? Due tipi distinti di approccio sono possibili:

Approccio "media-centrico"

In questo caso l'attenzione è focalizzata sulla componente essenziale del prodotto, il materiale audio-video. Le informazioni sono considerate accessorio d'arricchimento del materiale e la loro gestione non deve interferire con quella del materiale, anzi le due gestioni tendono a coincidere poiché le informazioni vanno ad accompagnare fisicamente audio e video.

Affinché questo approccio sia consistente occorre limitare accuratamente il tipo di informazioni da considerare.

Approccio "data-centrico"

L'approccio diametralmente opposto focalizza l'attenzione sulla gestione dei dati e delle informazioni, che si intende integrare su basi più ampie possibili. In questo caso il materiale audio video in forma elettronica rientra come

caso particolare nella gestione integrata. Il punto chiave per mantenere i collegamenti tra informazioni e materiale è l'identificazione univoca di quest'ultimo, dopodiché la gestione delle informazioni può avvenire con tecniche informatiche più generali (non necessariamente realizzate solo per l'industria televisiva). Questo approccio risulta comunque più complesso, in quanto implica un cambio di paradigma più radicale, e richiede un'attenta analisi dei benefici economici e creativi rispetto agli investimenti necessari.

3.3 Un nuovo modo di concepire i "metadati"

In questo scenario il ruolo di ciò che rientra comunemente sotto il nome di "metadati", ovverosia il ruolo dell'informazione che permea e governa i processi di produzione di tutta l'impresa e che rappresenta, identifica e descrive gli oggetti di un determinato dominio di operazione, risulta di primaria importanza. Questo tipo di informazione non è più accessoria e supplementare rispetto al media ma diventa fondamentale e complementare, nel momento in cui si vogliano dare delle risposte efficaci ai motori evolutivi presenti attualmente nel mercato dell'audiovisivo.

La modellazione dell'informazione e la progettazione dei sistemi informativi diventano alcune delle attività critiche e centrali in tutti i progetti di questa nuova era. Il progresso in questo particolare campo si muove attraversando alcune fasi:

- Fase 1: Riconoscimento della pari dignità tra le entità concettuali (ad esempio concetti editoriali, descrittori d'archivio) e quelle fisiche (ad esempio supporti d'archivio, file, apparati). La mera attribuzione dei metadati al materiale audiovisivo si trasforma in relazione tra entità concettuali ed entità fisiche, aprendosi per ciò stesso ad infinite possibilità aggiuntive di rappresentazione dell'informazione presente nel dominio. Il problema che si deve affrontare si evolve quindi, a partire dall'esclusiva individuazione degli attributi di un sottoinsieme delle entità (quelle fisiche) fino a comprendere l'individuazione delle *relazioni* possibilmente esistenti tra tutte le entità.

- Fase 2: Processo di integrazione dei sistemi e dei processi: spostamento dalla concezione di sistemi monolitici e concentrati a quella di sistemi distribuiti. La forza di tale approccio consiste nell'associare alle strutture periferiche (per esempio la produzione, gli archivi la messa in onda) la gestione delle informazioni e dei processi relative alle entità di propria competenza e di progettare sistemi informativi distribuiti sia in senso gestionale che fisico. Analogamente, si assiste allo sviluppo di modelli di scambio di dati standard sia a livello Enterprise che business to business[Nota 4]. L'integrazione in questa fase è attuata preminentemente a livello di tecnologia di scambio (per esempio si adotta XML[Nota 5] come tecnologia di formattazione e strutturazione dei dati durante lo scambio).

- Fase 3: Processo di integrazione dei modelli: la distribuzione dei sistemi implica che, per raggiungere modelli produttivi efficienti rispetto ai nuovi scenari di mercato, sia necessaria una quantità ingente di scambi di informazioni durante le fasi di vita del prodotto. Questo aspetto, tuttavia, non è da intendersi come confinato esclusivamente alle tecnologie di scambio dei dati. La piena maturità si raggiunge con l'integrazione dei modelli di rappresentazione, vale a dire con la definizione di un modello integrato e condiviso della conoscenza a livello di impresa (Enterprise). La condivisione della semantica dei dati a livello di impresa, contribuisce ad abbassare sensibilmente lo sforzo di adattamento delle applicazioni e dei sistemi, e quindi di riflesso ad abbassare i tempi di produzione.

Giunti a questo punto, entrano in campo alcuni elementi molto importanti per la fattibilità e gestibilità a lungo termine delle nuove imprese dell'audiovisivo: gli standard.

Nota 4 - Con l'etichetta "business to business" (B2B) si intende genericamente lo scambio di beni e servizi tra aziende o organizzazioni commerciali. In contrapposizione, tra le altre, con "business to consumer" (B2C) con la quale si intende lo scambio di beni e servizi tra aziende o organizzazioni commerciali e utenti/clienti finali

Nota 5 - eXstensible Markup Language, è una versione semplificata dell'SGML (Standard Generalised Markup Language), nato per permettere agli utenti di definire documenti contenenti codici di markup personalizzati. Il consorzio W3C (World Wide Web Consortium (www.w3.org), è l'organismo che ha standardizzato l'XML nel 1998.

4. Gli standard: perché sono necessari?

4.1 Standard a più livelli

In generale l'adozione di uno standard è motivata dalla ricerca di maggiori efficienza ed intercambiabilità.

- Lo standard può essere applicato a livello d'interfaccia per lo scambio del prodotto (formato). Nel caso dei metadati si tratta di avere un riferimento comune per identificare le informazioni necessarie, o utili, attribuirvi un significato condiviso e scambiarle .

- Ulteriormente possono essere standardizzate delle componenti utilizzate nel corso della realizzazione di un prodotto. Come esempi si possono riportare l'adozione di una certa applicazione software oppure di un certo apparato, oppure di una loro parte. Dal termine dell'era pionieristica della televisione, i broadcaster hanno fatto sempre più ricorso ai costruttori per approvvigionarsi di quel che serve tecnicamente e, naturalmente, questo vale anche per il caso informatico.

- Infine, nel caso si stabilisse un notevole grado di omologazione per un certo tipo d'attività, si può arrivare all'adozione di un processo standard per un'intera catena produttiva. In questo caso i vari operatori dello stesso ambito d'attività, agirebbero di fatto nello stesso modo, o comunque molto similmente. Nel caso della televisione non vi è, al momento, una tendenza certa all'omologazione, sia a causa della differenza in dimensioni tra grandi televisioni nazionali e le emittenti locali, sia perché l'attività può essere indifferentemente svolta tutta all'interno di una stessa azienda oppure suddivisa tra aziende più specializzate (archivi, case di produzione, editori, semplici radiodiffusori), sia per il desiderio di distinguere la propria offerta dalle altre.

Nota 6 - I concetti di business-to-business e business/ enterprise sono di fatto concetti relativi e non assoluti in quanto dipendono dal livello di divisionalizzazione di una organizzazione. Se all'interno di un'impresa si ha una suddivisione gestionale spinta, per esempio tra la funzione di archivio e quella di produzione, tra queste due funzioni si instaura normalmente un processo classificabile sotto il concetto di B2B.

Nel paradigma moderno queste necessità sono tutte di primaria importanza, perciò il bisogno di standard è sempre più sentito a qualsiasi dei tre livelli.

5. Lo stato attuale dell'evoluzione

5.1 Verso un paradigma moderno

La dinamica delineata porta alla definizione di un paradigma moderno. Le informazioni i processi produttivi e la conoscenza formalizzati e integrati producono e rendono economicamente attuabili nuovi spazi di sviluppo per l'organizzazione:

- Arricchimento dei contenuti sui nuovi canali . Le informazioni editoriali chiave a proposito delle produzioni possono essere usate efficacemente nella messa a punto di servizi di alimentazione di contenuti di portali Internet o di applicazioni di Video on Demand.

- Nuovi servizi & interattività. L'evoluzione dai servizi tradizionali di broadcasting analogico alle piattaforme digitali integrate, con canali di ritorno per l'interattività, integra il prodotto tradizionale (audiovisivo) con contributi di arricchimento. L'attuabilità delle soluzioni per questi servizi dipende fortemente dal grado di maturità dell'integrazione delle informazioni e dei processi a livello d'impresa.

- "Business to business". L'abbattimento dei costi di accesso all'archivio e l'evoluzione verticale dei processi di produzione genera una possibilità di ritorno positivo dalla commercializzazione dei prodotti d'archivio. Normalmente questo ambito viene classificato con l'etichetta "B2B" ovvero business-to-business[Nota 6]. Il caso Rai Click (www.raiclick.it) è un esempio di messa a punto di una prima integrazione tra archivio/produzione/ commercializzazione, che nasce proprio dall'instaurazione del paradigma

moderno attraverso un processo analogo all'evoluzione delineata.

5.2 I lavori in corso

Durante gli ultimi anni all'interno della comunità internazionale dei broadcaster e produttori, l'evoluzione descritta ha preso decisamente piede. I risultati di questo fenomeno sono piuttosto evidenti: per esempio si sono definiti, nell'ambito di organizzazioni di standardizzazione quali ISO/IEC, SMPTE, EBU e altri, nuovi standard per i metadati (MPEG7[Nota 7], P_META[Nota 8]), per i formati file (MXF[Nota 9], GXF[Nota 10]) e per il trasporto di contenuti su interfacce pacchettizzate (SDTI[Nota 11]). Anche a livello della comunità di utilizzatori le esperienze delle singole organizzazioni risultano significative; in particolare la Rai ha condotto sperimentazioni e sviluppato sistemi tuttora in produzione basati sull'esportazione dei metadati di archivio verso aziende dell'area Internet[Nota 12].

L'universo degli standard e la miriade di attività presenti a tutt'oggi intorno a queste problematiche saranno presentati e trattati nelle prossime uscite di questa serie di articoli, ma saranno comunque inquadrati nello scenario qui delineato.

Nota 7 - MPEG-7 è uno standard ISO/IEC, formalmente denominato "Multimedia Content Description Interface", e fornisce un insieme di strumenti standardizzati per la descrizione dei contenuti multimediali,

Nota 8 - P_META è uno schema di metadati nato per lo scambio di contenuti tra organizzazioni. E' stato sviluppato all'interno di un progetto EBU (European Broadcasting Union).

Nota 9 - Material eXchange Format, è un formato file in corso di standardizzazione dall'SMPTE (Society of Motion Picture and Television Engineers). E' previsto dagli addetti ai lavori che questo formato sarà pervasivo nel mondo della futura produzione TV basata su Information Technology. Permette il trasporto di audio video e metadati sincronizzati.

Nota 10 - General eXchange Format, è un formato file usato principalmente dalla linea di prodotti Grass Valley Group, ha ottenuto lo status di standard dell'SMPTE.

Nota 11 - Serial Digital Transport Interface. Estensione dello standard SDI (Serial Digital Interface) per il trasporto di dati compressi sui canali digitali audio/video.

Nota 12 - Questi sviluppi sono stati in gran parte possibili grazie all'esperienza maturata nella progettazione e messa a punto del sistema Catalogo Multimediale Rai.

Metadati e Modellazione

Laurent **Boch**,
Alberto **Messina**

Inizia con questo numero un appuntamento fisso, a cura dell'Unità Organizzativa Produzione del Centro Ricerche e Innovazione Tecnologica della Rai, dedicato alle problematiche relative alla definizione e all'uso dei *metadati* in contesti di produzione e distribuzione di contenuti audiovisivi e ai temi riguardanti la *modellazione* di tali contesti.

Il Centro Ricerche ha maturato negli ultimi anni una profonda esperienza in questi ambiti, in quanto coinvolto nella progettazione del Catalogo Multimediale Rai, di cui ha curato, tra gli altri, gli aspetti di modellazione dati e di analisi del contesto. Questa esperienza è stata ulteriormente consolidata grazie alla partecipazione attiva sia a progetti di standardizzazione internazionale, dedicati alla definizione di strutture di metadati, quali il progetto EBU denominato P/META, o alla modellizzazione dell'ambiente di produzione televisiva, quali il progetto EBU P/FTP, sia allo sviluppo di sistemi, tuttora in produzione, basati sull'utilizzo di tali metadati in contesti intra e interaziendali.

In questo spazio verranno ospitati sia articoli di carattere teorico e generale, sia scritti relativi più strettamente alle attività di standardizzazione e di sviluppo di sistemi.

L'articolo che segue, di carattere generale, è dedicato ad una analisi di scenario riguardo alla gestione dell'informazione in una moderna impresa di produzione e diffusione di contenuti multimediali. È da considerarsi quale introduzione ai futuri articoli e porta argomenti che giustificano la decisione di aprire questo spazio fisso.

Che cosa è, come funziona: Le origini del video digitale (La raccomandazione ITU-R BT.601)

ing. Marzio Barbero e
ing. Natasha Shpuza

1. Premessa

Alla base della rapida evoluzione dei sistemi di codifica digitale dell'informazione video sono le specifiche relative al campionamento e alla codifica individuate dalla Raccomandazione ITU-R BT.601. L'articolo [1] pubblicato nel 1982 da Elettronica e Telecomunicazioni forniva una completa e competente analisi della nuova raccomandazione: la prima versione risale infatti a tale anno. Questa scheda intende riproporre brevemente le caratteristiche principali di queste specifiche poiché esse spesso hanno implicazioni rilevanti nella definizione dei sistemi di compressione del segnale video e sulle tecniche di editing. La definizione della Rac. ITU-R BT.601 è l'evento alla base dello sviluppo e alla diffusione degli apparati di ripresa, di editing e di distribuzione e questa evoluzione è uno dei temi ricorrenti delle schede pubblicate in questa sezione "Che cosa è, come funziona" di Elettronica e Telecomunicazioni.

2. Cenni storici

Negli anni '70, le tecnologie numeriche si erano sufficientemente sviluppate da consentire all'industria di produrre i primi apparati digitali in grado di operare su segnali video di qualità adatta alla realizzazione di prodotti televisivi: nel 1981 l'UER e SMPTE riuscirono a definire i parametri essenziali per raggiungere un accordo a livello mondiale. In particolare i ricercatori della Rai contribuirono, così come quelli dei maggiori laboratori e centri di ricerca degli enti televisivi, alla sperimentazione tecnica, nell'ambito dei gruppi di lavoro UER e del ITU, necessaria alla definizione di tali parametri.

La prima versione delle specifiche si limitava alla definizione dei parametri relativi al primo membro di una "famiglia estensibile di standard per la codifica digitale compatibile" basata sull'adozione di una frequenza di campionamento unica, pari a 13,5 MHz, per i formati di immagine 4:3 a 625 righe/50 Hz e 525 righe/60 Hz. La versione attuale, la quinta del 1995 [2], specifica anche i parametri per la codifica nel caso di formato d'immagine 16:9, utilizzando la frequenza di campionamento di 13,5 MHz oppure 18 MHz, quando è richiesto un proporzionale aumento della risoluzione orizzontale.

3. Motivazioni alla base dello standard

La decisione di passare da una codifica del segnale video di tipo analogico (NTSC, PAL, SECAM) ad una codifica di tipo numerico ha varie motivazioni:

- Consentire l'uso di reti digitali per trasportare il segnale: i collegamenti numerici si possono considerare "trasparenti" al tipo di informazione trasportata (ad esempio: dati, audio, video) e al formato di codifica

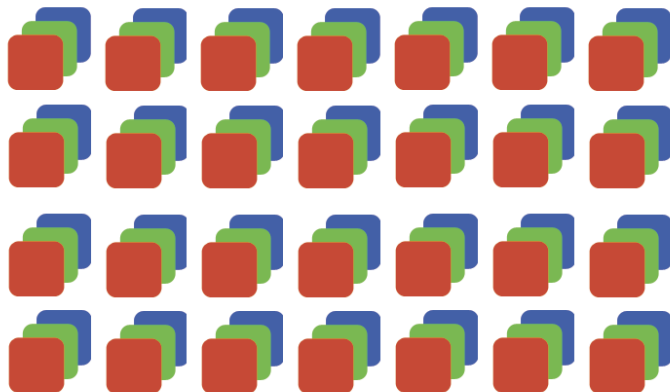



Fig. 1 - Nel caso di struttura di campionamento 4:4:4, le tre componenti sono campionate alla stessa frequenza e i tre campioni corrispondenti allo stesso elemento di immagine sono co-posizionati: ad esempio, nel caso di RGB i campioni rosso, verde e blu hanno lo stesso reticolo di campionamento.

dell'informazione (ad esempio, nel caso del video: formati a componenti o compositi, a 625 o 525 righe)

- Utilizzare apparati in grado di elaborare l'informazione video in modo da consentire l'ideazione di programmi e l'utilizzazione di linguaggi tecnico-espressivi (ad esempio effetti speciali e studio virtuale) assolutamente non pensabili nel caso si fosse continuato ad operare nel dominio analogico.

La raccomandazione è stata definita con lo scopo di avere la maggior parte dei parametri in comune nel caso dei formati a 625 e 525 righe al fine di consentire economie di scala nella realizzazione degli apparati e di facilitare lo scambio internazionale dei programmi. In particolare si è scelto di avere frequenze di campionamento uguali (per formati a 525 e 625 righe) e una codifica basata su tre componenti.

4. La famiglia estensibile di standard compatibili per la codifica digitale

4.1 4:4:4 e 4:2:2

I segnali possono essere sotto forma RGB, cioè corrispondenti ai tre colori primari (rosso, verde e blu, ovvero RGB *red-green-blue*). In questo caso tutte le tre componenti sono campionate alla stessa frequenza: i membri della famiglia che rispondono a queste caratteristiche sono denominati 4:4:4. In questo caso la struttura di campionamento prevede che i campioni relativi alle tre componenti siano co-posizionati (figura 1).

A partire dai segnali elettrici precorretti di gamma E'_R, E'_G e E'_B, all'uscita dei sensori della telecamera è possibile ottenere come loro combinazione lineare altri tre segnali digitali denominati Y (luminanza), C_R e C_B (differenza-colore), anche in questo caso la struttura è 4:4:4, cioè le tre componenti sono caratterizzate dalla stessa frequenza di campionamento e sono co-posizionati.

A partire dalle componenti Y, C_R e C_B nel formato 4:4:4 è possibile ottenere il formato 4:2:2, caratterizzato dal fatto che le due componenti differenza-colore sono limitate in banda e campionate a metà della frequenza con cui è campionata la luminanza.

Ovviamente la limitazione di banda implica che i formati 4:2:2 siano meno adatti ad applicazioni in cui si richieda una elevata qualità anche in presenza di numerose e complesse elaborazioni del segnale video.

D'altro canto la scelta di prevedere il formato 4:2:2 è essenzialmente economica: si riduce del 25% il numero di campioni da memorizzare o trasmettere. Inoltre sia i sistemi per la diffusione, sia quelli analogici (PAL, SECAM, NTSC) che quelli digitali (DVB, DVD) sono caratterizzati da forte limitazione di banda delle componenti di colore e quindi non vi sono degradamenti nel passaggio dal formato professionale a quello di diffusione.

Quando vennero definiti i parametri di codifica, molta attenzione fu fatta sulla scelta della struttura di campionamento e sulla banda necessaria per le componenti, per assicurare buone prestazioni anche nel caso di operazioni di postproduzione complesse, quali il chroma-key o intarsio [1]. Particolare cura è posta nelle specifiche per i filtri a cui sottoporre i segnali Y o R,G e B (figura 2) e i segnali differenza-colore sia nel caso siano ottenuti direttamente

a partire dai segnali analogici (figura 3), sia a partire dai segnali digitali mediante filtri numerici (figura 4).

Nel caso di struttura di campionamento 4:2:2 ciascuna coppia relativa ai due segnali differenze-colore deve essere co-posizionata spazialmente con i campioni di posizione dispari della luminanza (figura 5).

Fig. 2 - Specifiche per un filtro per i segnali di luminanza o RGB quando si utilizza la frequenza di campionamento 13,5 MHz: maschera per la caratteristica perdita d'inserzione in funzione della frequenza.

4.2 La quantizzazione

Le parole digitali che rappresentano i valori dei campioni possono essere a 8 bit o a 10 bit, gli otto bit più significativi sono la parte intera, mentre i restanti due bit, se presenti, sono da considerare parte frazionaria e se non sono indicati sono da supporre uguali alle cifre binarie 00

Fig. 3 - Specifiche per un filtro per i segnali di differenza-colore quando si utilizza la frequenza di campionamento 6,75 MHz: maschera per la caratteristica perdita d'inserzione in funzione della frequenza.

Con 8 bit sono disponibili 256 livelli di quantizzazione equispaziati (da 0000 0000 a 1111 1111 in notazione binaria, o da 0 a 255 in notazione decimale). Le configurazioni 0 e 255 sono riservate per la sincronizzazione, mentre quelle da 1 a 254 sono destinate al video.

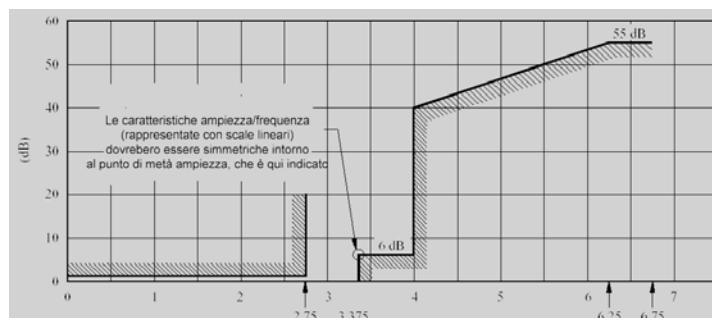



Fig. 4 - Specifiche per un filtro digitale per la conversione dalla frequenza di campionamento dei segnali differenza-colore per passare da 4:4:4 a 4:2:2: maschera per la caratteristica perdita d'inserzione in funzione della frequenza.

Per consentire opportuni margini operativi (ad esempio per evitare che calcoli per realizzare filtri digitali generino valori video non rappresentabili correttamente) l'informazione di luminanza occupa solo 220 dei livelli disponibili: il nero corrisponde al livello 16 e il bianco al livello 235. Analogamente i segnali differenze di colore devono occupare solo 225 livelli e il valore 0 di ciascun segnale differenza di colore deve corrispondere al livello digitale 128.

Fig. 5 - Nel caso di struttura di campionamento 4:2:2, il numero di campioni relativi alle componenti differenza-colore è metà rispetto ai campioni di luminanza. Entrambi i campioni differenza-colore sono co-posizionati con i campioni di ordine dispari relativi alla luminanza.

Fig. 6 - Una ripresa può essere effettuata con formato d'immagine 16:9 (a) oppure 4:3 (c), in entrambi i casi può essere rappresentata da 720 campioni in orizzontale e 576 righe. Per visualizzare correttamente le immagini occorre avere diplay opportuni e adottare la corretta deflessione (nel caso di display analogico) o interpolazione (nel caso di display digitale). L'immagine 16:9 riprodotta ha ovviamente una definizione orizzontale inferiore. Le figure (b) e (d) rappresentano le immagini nel caso in cui si utilizzi un display a pixel quadrato, senza interpolazione, per visualizzare rispettivamente (a) e (c): vi è una deformazione spaziale dell'immagine particolarmente evidente nel caso del formato 16:9 (b), ma presente anche nell'altro caso (d) poiché 720 e 576 non sono in rapporto 4:3.

4.3 I membri della famiglia

La tabella 1 riporta i parametri per il membro 4:2:2 e frequenza di campionamento 13,5 MHz della famiglia di standard. Questi parametri si applicano sia nel caso in cui il formato d'immagine sia 4:3 sia nel caso in cui sia 16:9 (figura 6).

La scelta della stessa frequenza di campionamento per i sistemi a 625 righe e quelli a 525 righe facilita la realizzazione dei sistemi di instradamento e commutazione del segnale televisivo numerico negli studi e centri di produzione televisivi: il flusso binario complessivo è pari, in entrambi i casi, a 270 Mbps (l'interfaccia di interconnessione è specificata da una successiva raccomandazione [3] dove i campioni di luminanza e di differenza-colore sono multiplati e a ciascun campione corrispondono 10 bit).

Si osservi che il numero di campioni costituenti la riga attiva è uguale per i sistemi a 625 righe e quelli a 525 righe (720 campioni Y), malgrado differiscano le durate totali delle righe (sono rispettivamente 864 e 858). Questa scelta ha consentito di facilitare la struttura dei sistemi in grado di elaborare o memorizzare il segnale video digitale. Non solo la quantità di dati per riga attiva è costante, ma anche la quantità di dati utili (parte attiva del segnale video) complessiva è costante: infatti, anche se questo parametro non è indicato nella [2], il numero di righe attive normalmente elaborato o memorizzato è rispettivamente 576 nel caso di sistemi a 25 quadri al secondo e 480 nel caso di quelli a 30 quadri al secondo e pertanto il numero totale di elementi d'immagine risulta essere in entrambi i casi 10 368 000 al secondo.

Nel caso del membro 4:4:4 (utilizzabile sia per R,G, B che per Y, C_R e C_B) valgono parametri del tutto simili a quelli indicati nella tabella precedentemente citata, tranne che per tutte le tre componenti valgono la stessa frequenza di campionamento (13,5 MHz) e lo stesso numero di campioni per riga totale e per riga attiva (720 per ciascuna componente).

Parametri	Sistemi a 525 righe e 60 semiquadri	Sistemi a 625 righe e 50 semiquadri
Segnali codificati Y, C_R , C_B	Questi segnali sono ottenuti a partire dai segnali precorretti di gamma, vale a dire E'_Y, E'_R-E'_Y, E'_B-E'_Y	
Numero di campioni per l'intera riga - segnale di luminanza (Y) - ciascun segnale differenza-colore (C_R, C_B)	 858 429	 864 432
Struttura di campionamento	Ortogonale, ripetitiva a livello di riga, semiquadro e quadro. I campioni C_R, C_B sono coposizionati rispetto ai campioni dispari (1°, 3°, 5° …) in ciascuna riga	
Frequenza di campionamento - segnale di luminanza - ciascun segnale differenza di colore	 13,5 MHz 6,75 MHz La tolleranza per la frequenza di campionamento deve coincidere con la tolleranza per la frequenza di riga dello standard televisivo a colori corrispondente	
Formato di codifica	PCM uniformemente quantizzato, 8 (opzionalmente 10) bit per campione, per il segnale di luminanza e per ciascuno dei segnali differenza di colore	
Numero di campioni per riga attiva digitale - segnale di luminanza - ciascun segnale differenza di colore	 720 360	
Relazione temporale orizzontale analogico-digitale - dalla fine della riga attiva digitale a O_H	16 periodi di clock per la luminanza	12 periodi di clock per la luminanza
Corrispondenza tra i livelli del segnale video e i livelli di quantizzazione - estensione della scala di quantizzazione - segnale di luminanza - ciascun segnale differenza di colore	Da 0 a 255 220 livelli di quantizzazione con il livello del nero corrispondente a 16 e il livello di picco del bianco corrispondente a 235. Il livello del segnale può occasionalmente superare 235 225 livelli di quantizzazione centrato rispetto alla scala di quantizzazione con il livello zero corrispondente a 128	
Utilizzo delle parole di codice	Le parole di codice corrispondenti ai valori numerici 0 e 255 sono usati esclusivamente per la sincronizzazione. I valori da 1 a 254 sono disponibili per il video	

Tab. 1 - Valori dei parametri di codifica per i membri 4:2:2, 13,5 MHz della famiglia. Sono utilizzati sia nel caso di televisione digitale con formato d'immagine 4:3, che nel caso di utilizzo con formato di visualizzazione 16:9.

Nel caso in cui il formato d'immagine sia 16:9 e si voglia conservare un'elevata risoluzione orizzontale, cioè non sia accettabile la perdita di risoluzione passando da 4:3 a 16:9 quando si utilizza uno dei due precedenti membri della famiglia, si deve utilizzare uno dei due membri della famiglia caratterizzati dalla frequenza di campionamento pari a 18 MHz (in tabella 2 sono riportati i parametri relativi al numero di campioni per riga).

Parametri	Sistemi a 525 righe e 60 semiquadri	Sistemi a 625 righe e 50 semiquadri
Numero di campioni per l'intera riga segnale di luminanza (Y) ciascun segnale differenza-colore (C_R, C_B)	 1144 572	 1152 576
Frequenza di campionamento segnale di luminanza ciascun segnale differenza-collore	 18 MHz 9 MHz	
Numero di campioni per riga attiva digitale segnale di luminanza ciascun segnale differenza di colore	 960 480	

Tab. 2 - In questa tabella sono riportati solo I valori dei parametri per i membri 4:2:2, 18 MHz della famiglia che differiscono da quelli in tabella 1.

5. I problemi di conversione

Inizialmente il video digitale veniva acquisito e utilizzato soprattutto nel formato 4:2:2, con formato di immagine 4:3. L'evoluzione delle tecnologie digitali in campo video è stata molto rapida e oggigiorno è possibile produrre e post-produrre a costi limitati e con qualità elevata (in molti casi, sia per effetti speciali che per correzioni cromatiche è necessario operare in 4:4:4) ed è sempre più diffuso il formato d'immagine 16:9 (supportato dal DVB, dal DVD e fruibile mediante i nuovi televisori a grande schermo). Diventano sempre più frequenti le conversioni dei vari membri della famiglia, fra di loro e verso i formati di ripresa e di editing non lineare semiprofessionale e consumer.

Limitandosi ai soli ambiti di applicazione della Rac. ITU-R BT.601, si può comunque far cenno ai seguenti problemi che, solo se noti e tenuti in considerazione, consentono di sfruttare al meglio le caratteristiche di qualità dei segnali video digitali:

- La corrispondenza fra i livelli compresi fra nero e bianco è in ambito video, come si è visto precedentemente, è ristretta alle rappresentazioni numeriche fra 16 e 235, ma questa limitazione non è spesso presa in considerazione dai sistemi di grafica elettronica (in genere di derivazione informatica, per i quali al nero corrisponde 0 e al bianco corrisponde 255).

- L'insieme dei colori rappresentabili nella forma Y, C_R e C_B è più ampio di quello rappresentabile con R, G e B: ciò può avere delle implicazioni sulla generazione ed elaborazione di immagini in ambito digitale. E' opportuno limitare i segnali Y, C_R e C_B, sacrificando eventualmente la saturazione, prima che, convertiti in R, G e B, possano dare origine a degradamenti non desiderati e percettibili su luminanza e tonalità di colore.

- Nel caso di postproduzioni complesse le operazioni di interpolazione e decimazione necessarie per passare da 4:2:2 a 4:4:4 possono dare origine a degradamenti percepibili. D'altro canto è possibile che si verifichino conversioni multiple di questo tipo, poiché i sistemi di videoregistrazione professionali sono in genere di tipo 4:2:2 e quelli di derivazione consumer riducono ulteriormente la banda dei segnali differenza-colore.

- Il rapporto fra il numero di campioni attivi per riga (720) e quello delle righe attive (576) non corrisponde a 4:3, ovvero 1,33, bensì a 1,25: si dice che il pixel non è "quadrato", a differenza di quanto avviene nel caso di sistemi fotografici o di grafica elettronica. Le conseguenze di questo fatto sono che un'immagine 4:3 video risulta lievemente distorta sullo schermo di un computer, come percepibile confrontando le figure 6c e 6d, oppure che immagini create in ambito informatico appaiono deformate sullo schermo televisivo. In particolare le forme geometriche (cerchi e quadrati) possono apparire visibilmente deformate (ellissi e rettangoli). Quando si produce grafica o titoli, per poi convertirli in quadri video, utilizzando sistemi e programmi grafici realizzati per applicazioni informatiche, è quindi buona norma realizzarli in formato 768 x 576 (rapporto 4:3) e trasformarli nel formato 720 x 576 al momento della loro conversione in immagini video digitale secondo la Rac. ITU-R BT.601.

Bibliografia

1. G.F. Barbieri: "Codifica Numerica del segnale video: standard per gli studi televisivi", Elettronica e Telecomunicazioni, n. 2, 1982, pp. 42-52

2. Recommendation ITU-R BT.601-5: "Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios" (1982-1986-1990-1992-1994-1995)

3. Recommendation ITU-R BT.656-4: "Interfaces for Digital Component Video Signals in 525-line and 625-line Television Systems operating at the 4:2:2 Level of Recommendation ITU-R BT.601-5 (Part A)", (1986-1992-1994-1995-1998)

Che cosa è, come funziona

Marzio Barbero
Natasha Shpuza

Questa sezione è costituita da schede sulle tecnologie alla base dei sistemi di telecomunicazioni ed in particolare della televisione. Le schede pubblicate in questo numero sono dedicate all'evoluzione delle tecniche digitali per la codifica video.

La definizione dello standard di codifica video con qualità adatta alla radiodiffusione (*broadcasting*) è del 1982. Nel 1990 le tecniche di riduzione e compressione della ridondanza per segnali a definizione standard ed ad alta definizione dimostrano la possibilità di portare la TV digitale all'utente finale, favorendo infine la definizione dello standard MPEG-2 nel 1995 e successivamente l'introduzione del DVB e del DVD. In questi mesi è in fase di definizione tecnica finale il nuovo standard per la compressione video (denominato AVC o H.264), le cui caratteristiche sono oggetto dell'ultima scheda.

Per quanto riguarda invece l'evoluzione in corso delle tecnologie per i display piatti, oggetto delle schede del numero di agosto 2003 di Elettronica e Telecomunicazioni, è da segnalare l'introduzione, nel mese di aprile, di un un primo prodotto commerciale destinato alla diffusione in grandi volumi che utilizza un display OLED a matrice attiva (gamma completa di colori, angolo di visione 165°) da 2,2" a corredo di un apparato fotografico della Kodak (presso i cui laboratori furono sviluppati i primi OLED).

Relativamente alle tecnologie per i dischi ottici, oggetto delle schede del numero di dicembre 2002, sono interessanti le novità relative all'introduzione di prodotti basati sul Blu-Ray Disc. Nel dicembre 2002 è stato raggiunto un accordo tra Nichia (presso i cui laboratori è stato dimostrato nel 1995 il primo laser blu-violetto) e Sony con l'obiettivo di condividere know-how e brevetti al fine di sviluppare le tecnologie per la produzione di massa del dispositivo laser entro la primavera di quest'anno. In effetti è stato annunciato l'inizio della commercializzazione, a partire dal mese di aprile in Giappone, di un registratore completo di sintonizzatore per la ricezione della TV satellitare. Su un disco è in grado di registrare 23 GB, pari a circa 2 ore di TV ad alta definizione (a 24 Mbps) e fino ad un massimo di 12 ore di TV a definizione standard nella modalità LR (long time mode) a 4 Mbps. I prezzi indicati sono 450 000 ¥, circa 3460 €, per l'apparato e 3500 ¥, circa 27 € per il disco riscrivibile. Nel mese di aprile, alla mostra annuale NAB (*National Association of Broadcasters*) di Las Vegas è stata presentata dalla Sony una linea di prodotti professionali per ENG (*Electronic News Gathering*), EFP (*Electronic Field Production*) e per applicazioni di cinema elettronico: della linea fanno parte deck per l'editing e camcorder che operano in formato IMX e DVCAM e utilizzano i blu-ray disc riscrivibili da 23 GB.

Fig. - immagini del registratore blu-ray disc e del disco riscrivibile (*cartridge*) tratti da www.sony.jp

Che cosa è, come funziona: Utensili per la compressione video

ing. Marzio Barbero e
ing. Natasha Shpuza

1. Premessa

Si è visto, nella scheda precedente, che il segnale video digitale codificato secondo la Rac. ITU-R BT.601 è caratterizzato da un numero di campioni elevato: vi sono 10 368 000 elementi d'immagine (la sola porzione attiva, senza considerare i sincronismi) al secondo. Considerando il formato 4:2:2 e una codifica a 8 bit al campione per la luminanza e altrettanti per le due componenti di crominanza, il bit rate associato all'informazione video è quindi circa 166 Mbps.

Negli anni '80 si realizzarono i sistemi atti a manipolare e registrare un tale flusso di dati, ma era evidente la necessità di sviluppare tecniche per ridurre il bit-rate necessario al trasporto, poiché le capacità dei canali (in particolare i ponti radio numerici) disponibili non consentivano il trasferimento dei segnali sotto forma digitale tra i luoghi di ripresa, produzione e postproduzione.

Per ridurre il bit-rate associato al video, senza compromettere significativamente la qualità delle immagini si sviluppò e si ottimizzò un insieme di algoritmi o utensili (*tool-kit*) per comprimere il segnale video e consentirne il transito per i canali disponibili.

2. Cenni storici

I ponti radio disponibili negli anni '80 per la rete di contribuzione (collegamenti fra studi e centri di produzione) e di distribuzione primaria (per trasferire il segnale ai centri di diffusione terrestre) avevano una capacità di 34 Mbps (in Europa) e di 45 Mbps (negli Stati Uniti). Era quindi necessario definire uno standard per comprimere il segnale di almeno un fattore 6 mantenendo una qualità dell'informazione video tale da consentire eventuali operazioni di post-produzione.

Per ottenere livelli elevati sia per il fattore di compressione che per la qualità occorre utilizzare contemporaneamente tecniche (sviluppate a partire dagli anni '50) che sfruttano la ridondanza statistica dei dati, tecniche basate sulla ridondanza spaziale (la trasformata DCT era stata proposta per le immagini nel 1974 [1]) e sulla ridondanza temporale (compensazione del movimento), ottimizzando la scelta di algoritmi e parametri di codifica in funzione delle caratteristiche psicovisive umane.

Per questa opera di definizione e ottimizzazione degli algoritmi fu molto importante il contributo dei centri di ricerca e sviluppo dei radiodiffusori e dei produttori di apparati per telecomunicazioni europei, in particolare di quelli partecipanti al progetto europeo Eureka 256, fra cui il Centro Ricerche della Rai. Tale progetto mise a punto un sistema in grado di operare anche sul segnale ad alta definizione, con un fattore di compressione superiore a 10, e culminò nel 1990 con la realizzazione degli apparati che consentirono la trasmissione da parte della Rai delle partite del campionato mondiale di calcio. I risultati tecnologici ottenuti furono oggetto di numerose pubblicazioni [2-5] e riconoscimenti a livello internazionale (figura 1).

Fig. 1 - La copertina e l'indice del numero speciale di Elettronica e Telecomunicazioni del novembre 1990 dedicato alla trasmissione di alcune partite del campionato mondiale di calcio Italia '90 in alta definizione e codificate con il sistema di riduzione della ridondanza messo a punto nell'ambito del progetto Eureka 256.
Pagina del numero 2 del 1991 "Assegnazione del Montreaux Achievement Gold Medal all'ing. Marzio Barbero del Centro Ricerche della Rai (Montreux, 13-18 giugno 1991) ".

Fig. 2 - Uno dei codificatori HDTV utilizzati per Italia '90. Il codificatore è racchiuso in un sottotelaio da 19" di larghezza e 6 unità di altezza. Il consumo è pari a circa 200 W.

Per quanto riguarda la tv a definizione convenzionale, nel 1992 fu emesso lo standard europeo ETS 300 174 [6] a cui corrisponde a livello mondiale la Rac. ITU-T J.81 [7].

A partire da metà anni '90 si è poi rapidamente sviluppata la tecnologia, in particolare quella legata al sistema MPEG-2 oggetto della scheda successiva, e ciò ha comportato ulteriori miglioramenti dal punto di vista degli algoritmi, ma soprattutto dal punto di vista delle possibilità di integrazione, passando da apparati delle dimensioni di un *rack* (figura 2) a circuiti integrati caratterizzati da minimo ingombro e consumo: un dispositivo recentemente messo in commercio è costituito da un singolo *package* plastico a 273 piedini, dimensioni 15x15 mm, consumo 330 mW, è in grado di co/decodificare segnali audio/video MPEG-2.

Fig. 3 - Albero binario costruito a partire dalla statistica dei simboli presenti nella stringa di esempio.

3. Ridondanza statistica e compressione *lossless*

Per ridurre il bit-rate associato ad un segnale digitale si può sfruttare la ridondanza di tipo statistico presente nel flusso di dati, utilizzando tecniche sviluppate a partire dagli anni '50 per ridurre il numero di bit associato a testi o a dati di tipo informatico.

3.1 Codice di Huffman

Uno dei metodi più conosciuti è quello proposto dal matematico D.A. Huffman nel 1952.

Questo algoritmo è applicabile quando la sorgente emette simboli caratterizzati da una probabilità non uniforme.

Per capire come funziona, utilizziamo un semplice esempio basato su un breve testo:

sopra_la_panca_la_capra_campa._sotto_ la_panca_la_capra_crepa.

E' costituito da un totale di 61 caratteri in cui si individuano 13 simboli differenti (comprendendo fra i simboli anche lo spazio e il punto).

Per codificare ciascun simbolo in binario si potrebbero utilizzare parole costituite da 4 bit (con quattro bit le configurazioni possibili sono 2^4 cioè 16) e quindi l'intera stringa sarebbe rappresentabile da 244 bit in totale.

La codifica secondo Huffman permette di ridurre il numero di bit totali associando a ciascun simbolo una parola binaria di lunghezza non fissa, tale per cui ai simboli più probabili vengano associate parole più corte, ai simboli meno probabili parole più lunghe.

L'algoritmo funziona in questo modo:

- Analizza il numero di ricorrenze di ciascun simbolo (nell'esempio *a* ricorre 16 volte, lo spazio _ 11 volte, mentre *m* ed *e* compaiono una sola volta)
- Accomuna i due elementi meno frequenti in un sottoinsieme somma (nell'esempio *m* ed *e*) e li distingue associando, ad esempio, *0* ad *m* e *1* ad *e*.

Tab. 1 - Tabella di assegnazione dei codici binari in base all'albero di figura 1.

simbolo	numero di occorrenze	probabilità di occorrenza	parola di codice
a	16	0,262	10
_	11	0,180	00
p	7	0,115	010
c	6	0,098	1111
l	4	0,066	0111
r	4	0,066	0110
o	3	0,049	11101
n	2	0,033	11001
s	2	0,033	11000
t	2	0,033	11011
.	2	0,033	11100
e	1	0,016	110101
m	1	0,016	110100

- Ripete iterativamente il processo con i due sottoinsiemi meno frequenti, usando lo stesso procedimento e considerando un tutt'unico il sottoinsieme costituito da *m* ed *e*, caratterizzato da una probabilità di occorrenza pari alla somma delle probabilità associate ai singoli *m* ed *e*.

- Si crea così un albero (figura 3) costituito da una serie di ramificazioni binarie, in cui le foglie costituite dai simboli più rari sono più lontane dalla radice e sono identificate da un codice binario più lungo.

- seguendo il percorso dalla radice fino alla foglia, si determina il codice assegnato a ciascun simbolo (tabella 1).

Il decodificatore, seguendo il percorso indicato dai bit che esamina in sequenza è in grado di individuare univocamente (un codice di Huffman non è mai prefisso di un altro) la foglia, ovvero il simbolo relativo a ciascuna parola, anche se le parole sono a lunghezza variabile (VLC, *Variable Lenght Code*).

Tornando all'esempio, la stringa di caratteri "*sopra*" diventa, codificata in binario *11000*11101*010*0110*10.*

Nel caso di questa stringa, un codice a lunghezza fissa, assegnando 4 bit a simbolo, avrebbe richiesto 20 bit, mentre con la codifica VLC ne sono sufficienti 19. Un guadagno significativo si ha nel caso della stringa "_*la*_", che viene codificata con *00*0111*10*00 ovvero con, mediamente, 2,5 bit per carattere.

Nel complesso, la frase richiede 198 bit, anziché i 244 richiesti da una codifica a lunghezza fissa, con un risparmio dell'ordine del 19% e senza perdita di informazione (codifica *lossless*).

L'esempio è semplice e riduttivo: in genere i fattori di compressione ottenibili sono più elevati partendo da testi contenenti caratteri ASCII, codificati con parole a lunghezza fissa di 7 bit.

Il metodo funziona se codificatore e decodificatore utilizzano lo stesso albero, ovvero la stessa tabella, e ciò può essere ottenuto o adottando una tabella di assegnazione fissa, oppure inviando (nel caso di trasmissione) o memorizzando (nel caso di un file) la tabella prima dei dati compressi. E' anche possibile utilizzare l'algoritmo in modo adattativo, ovvero si parte da una tabella che viene aggiornata parallelamente dal codificatore e dal decodificatore in funzione dei simboli via via trasmessi.

3.2 Algoritmo LZW

Un ulteriore miglioramento di efficienza può essere ottenuto considerando, anziché i singoli simboli emessi dalla sorgente, insiemi di simboli. Tornando all'esempio precedente, un notevole guadagno è ottenibile utilizzando una singola parola di codice per rappresentare "_*la*_" che compare 4 volte e "*pra*" che compare 3 volte.

Questa tecnica è alla base dell'algoritmo noto come LZW (da Jacob Ziv e Abraham Lampel che pubblicarono due articoli nel 1977 e 1978 e Terry Welch che propose una modifica alle loro proposte nel 1984). Codici basati su LZW sono utilizzati, ad esempio, per la compressione di immagini (in formato GIF, *Graphics Interchange Format).* A seguito di controversie sui brevetti alla base della tecnica LZW e quindi GIF, è stato successivamente definito il formato PNG (*Portable Network Graphics*).

3.3 Run-Length Encoding (RLE)

Questo tipo di codifica si basa sulla ripetizione all'interno del messaggio di uno stesso simbolo. In questo caso è possibile, ad esempio codificare l'occorrenza di n simboli uguali con due byte, il primo, denominato *run count*, indica il numero di ripetizioni ed il secondo, denominato *run value*, indica il valore (ad esempio il codice ascii di un carattere, o il livello di luminanza di un pixel).

Fig. 4 - Confronto fra le dimensioni relative di un'immagine SDTV (*Standard Definition TV*) da 720x576 elementi di immagine, quella CIF (*Common Intermediate Format*, 352x288), QCIF (*Quarter CIF., 176x144*), QQCIF (88x72).

4. Irrilevanza e compressione *lossy*

Le tecniche precedentemente descritte sono efficaci nel caso di compressione di immagini grafiche, ed in effetti sono alla base di tutti i formati utilizzati per le immagini per PC, soprattutto quando vi sono sequenze di pixel uguali fra loro o strutture ripetitive, ma non consentono di ottenere elevati fattori di compressione nel caso di immagini di tipo naturale.

Per ottenere fattori di compressione superiori è necessario accettare la perdita di informazione (sistemi *lossy*), ovviamente riducendo al minimo la percezione dei difetti introdotti.

4.1 Ridondanza spaziale

Per ridurre il bit-rate finale un primo approccio è quello di ridurre il numero di campioni dell'immagine da codificare. In figura 4 sono messe a confronto le dimensioni relative di un'immagine da 720 elementi di immagine per riga e 576 righe, come previsto dalla Rac. ITU-R BT.601, con il formato CIF, utilizzata in applicazioni di videoconferenza e in MPEG-1. Il formato QCIF è spesso utilizzato per il webcasting, mentre il QQCIF è impiegato per videotelefonia (UMTS). La riduzione della definizione spaziale, orizzontale e verticale, è senz'altro efficiente e può essere efficace per immagini relative al voltro umano, o a singoli oggetti o animali (un fiore o una farfalla), ma non è utilizzabile in campo radiodiffusivo, a causa della difficoltà nel riprodurre i dettagli.

Fig. 5 - Due quadri successivi (sola luminanza) di una sequenza video e immagine differenza fra le due.

4.2 Ridondanza temporale

Le sequenze video sono caratterizzate da un'altra forma di ridondanza, quella temporale. Nel sistema europeo vi sono 25 quadri (ciascuno composto da due semiquadri interlacciati) al secondo. La risoluzione temporale è generalmente ridotta nelle applicazioni menzionate precedentemente: nel formato CIF di fatto si elimina un semiquadro su due, ma spesso il numero di quadri è ulteriormente ridotto (ad esempio 12 o 6 quadri al secondo). Ovviamente anche in questo caso l'informazione eliminata non è irrilevante ed il degradamento è sensibile.

Un ulteriore modo per sfruttare la ridondanza temporale è quello di codificare le differenze fra quadri successivi, trasmettendo o memorizzando solo l'informazione che cambia, da un quadro a quello successivo. Dalla figura 5 appare evidente che spesso le differenze fra due quadri successivi sono piccole, molto prossime al valore 0: è quindi possibile sfruttare tecniche di riduzione della ridondanza statistica, sfruttando la distribuzione non uniforme delle ampiezze delle differenze (i campioni video di luminanza sono invece distribuiti uniformemente fra il nero e il bianco).

4.3 Compensazione del movimento

I campioni co-posizionati del quadro precedente possono quindi costituire una predizione molto buona per i campioni da codificare. Nel caso che gli oggetti ripresi si muovano all'interno della scena oppure la telecamera effettui movimenti orizzontali (*panning*) o verticali (*tilting*) è possibile ottenere una buona predizione se si riesce a stimare il movimento e a codificare sia la predizione che il vettore movimento che permette al decodificatore di individuare l'informazione già trasmessa (parte del quadro precedente) da cui si è ottenuta la predizione. Un'ottima predizione è possibile associando un vettore movimento a ciascun campione d'immagine, ma ciò richiede la trasmissione di un enorme numero di vettori; si potrebbe trasmettere un solo vettore movimento, globale per l'intera immagine, ma in tal caso sarebbe trascurabile il guadagno di predizione: il compromesso in genere consiste nell'organizzare l'immagine in blocchi (ad esempio di 8x8 o 16x16 campioni) ed individuare e trasmettere un solo vettore movimento per ciascun blocco (figura 6).

Fig.6 - Un oggetto si trova in due posizioni differenti in due quadri successivi: un vettore, caratterizzato dalle due componenti x e y può consentire di compensare la traslazione dei pixel corrispondenti ad un oggetto che nei 40 ms intercorsi fra i due quadri appare spostato spazialmente, nell'immagine.

4.4 Quantizzazione

Una riduzione del bit-rate è ottenibile quantizzando con una minor precisione i campioni (figura 7): la perdita di informazione (indicata come errore di quantizzazione o rumore di quantizzazione) è rilevante e proporzionale al numero di bit al campione risparmiati.

Fig.7 - la riduzione del numero di bit per campione dà origine all'incremento del rumore di quantizzazione: in questa immagine di sola luminanza si ha un effetto di solarizzazione crescente a partire da quella in alto a sinistra (8 bit) a quella in alto a destra (6 bit), in basso a sinistra (4 bit), in basso a destre (2 bit).

5. Una trasformazione di dominio

Di fatto tutti i metodi precedentemente descritti sono utilizzati nei sistemi di compressione, sia quello sviluppato in Eureka 256, sia in quelli MPEG.

Non vengono però applicati direttamente ai campioni video perché, come si è visto, ad elevati fattori di compressione corrisponderebbero altrettanti elevati fattori di distorsione e perdita in qualità dell'immagine. Prima di essere compressi i campioni video vengono raggruppati (in genere in blocchi 8x8) e trasformati.

La trasformazione normalmente adottata è la DCT (*Discrete Cosine Transform*). La DCT è un algoritmo matematico che può essere descritto in molti modi: moltiplicazione matriciale, rotazione d'asse in uno spazio a 64 dimensioni, FFT (*Fast Fourier Transform*) di un blocco accanto al suo blocco riflesso, blocco decomposto nelle sue funzioni basi.

Comunque la si consideri, lo scopo della Trasformata Coseno Discreta è quello di ottenere, a partire dal blocco di 8x8 campioni video (8 bit per campione), un altro blocco di 8x8 valori, (generalmente rappresentati da 11 o 12 bit): a questo punto del processo il numero di bit associato a ciascun blocco è aumentato, ma la distribuzione statistica dei valori è radicalmente modificata.

Le funzioni basi rappresentano le frequenze spaziali, dalla continua alle frequenze orizzontali, verticali e diagonali più elevate (figura 8).

Il sistema psicovisivo umano considera meno rilevanti le frequenze spaziali più elevate e quindi si può applicare di fatto una riduzione di banda: i coefficienti sono pesati dividendoli per costanti di peso diverso a seconda della posizione nella matrice (figura 9).

Nel caso ciò sia conveniente, i blocchi trasformati non sono quelli contenenti i campioni video, ma quelli ottenuti come differenza, a partire da valori di predizione determinati anche sfruttando la compensazione del movimento.

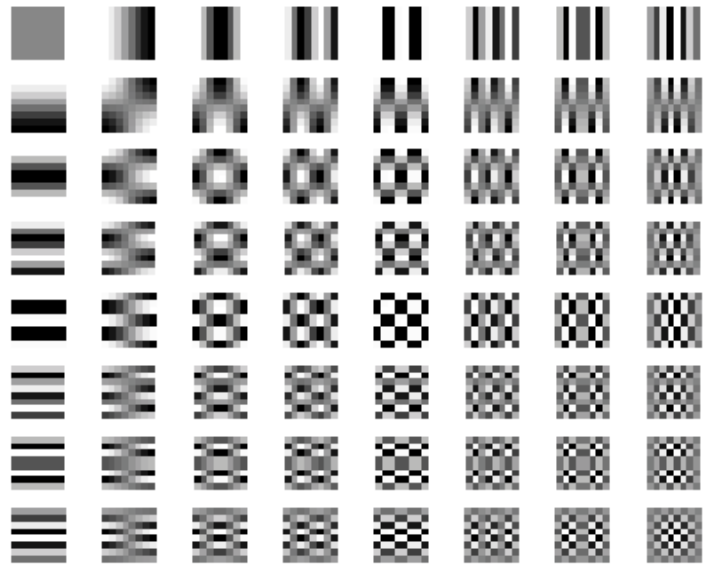



Fig.8 - Le 64 funzioni base nel caso della trasformata DCT 8x8. Qualsiasi blocco di campioni video può essere rappresentato da una combinazione di alcune delle funzioni opportunamente pesate. In genere è presente la prima funzione in alto a sinistra, che rappresenta la componente continua (il valore medio del blocco) più un certo numero di funzioni, con una predominanza di quelle corrispondenti alle frequenze spaziali più basse, cioè disposte nell'angolo in alto a sinistra.

Fig.9 - La matrice di pesatura è utilizzata per dividere i coefficienti per un valore legato al rumore di quantizzazione che il sistema psicovisivo è in grado di tollerare in funzione delle singole frequenze spaziali: il minimo di rumore è tollerato per la componente continua (cioè per i blocchi e, in definitiva, le immagini in cui ci sono lente variazioni di luminanza), mentre un maggiore errore di quantizzazione è accettabile per blocchi rumorosi o ricchi di dettagli. Questa è la matrice di pesatura adottata per MPEG-2.

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Fig.10 - I coefficienti moltiplicativi delle frequenze base sono trasmessi seguendo un ordine secondo un percorso a zig-zag: dapprima il coefficiente relativo alla componente continua e via via quelli relativi alle frequenze spaziali più elevate.

I coefficienti vengono riordinati, prima della trasmissione, secondo un percorso a zig-zag (figura 10) I coefficienti sono codificati mediante VLC, sfruttando così la statistica non uniforme delle loro ampiezze, mentre le lunghe sequenze di zeri che si vengono a creare possono essere codificate con RLE. Poiché in genere gli ultimi coefficienti sono nulli, essi non vengono trasmessi, ma l'ultimo *run-length* è, qualunque sia la sua lunghezza, sostituito da un unico simbolo, denominato EOB (*End of block*).

La lunghezza totale del blocco codificato è quindi variabile e dipende dalla configurazio-

Grazie alla ridondanza temporale, i coefficienti DCT, soprattutto quelli relativi alla componente continua e alle basse frequenze spaziali, hanno ampiezza inferiore a quelli che si otterrebbero trasformando direttamente i campioni video.

I 64 coefficienti moltiplicativi delle funzioni base assumono quindi valori prossimi allo zero e, anche grazie all'effetto della matrice di pesatura, sono spesso nulli quelli posizionati verso l'angolo inferiore a destra (frequenze spaziali più elevate).

Fig.11 - Se si suddivide un'immagine (la trattazione si riferisce alla luminanza, ma può essere estesa alle componenti di crominanza) in blocchi 8x8, questi hanno in genere caratteristiche diverse. Ad esempio il blocco in alto è tratto da una porzione del viso e si notano minori variazioni (i campioni sono più correlati), il blocco intermedio è tratto dalla siepe ed è caratterizzato da variazioni più ampie della luminanza (è una struttura casuale, simile al rumore), il blocco in basso è relativo alla collana e quindi rappresentativo di contorni e dettagli ben definiti.
Ciascun blocco, applicando la DCT, è rappresentabile come combinazione delle funzioni base.
Il numero di coefficienti generato è minore per il blocco superiore, il rumore di quantizzazione è più percepibile nel caso blocco inferiore rispetto a quello intermedio.

ne del blocco dei campioni video che lo ha originato.

I blocchi 8x8 presentano livelli di difficoltà differenti per la compressione (figura 11). I blocchi che danno origine ad un numero inferiore di bit sono quelli in cui vi è una forte correlazione, sono anche quelli su cui è più visibile l'eventuale rumore di quantizzazione. I blocchi che hanno una struttura più complessa, simile al rumore, sono quelli che danno origine ad un numero di coefficienti e, in ultima analisi, di bit superiore, ma sono anche quelli in cui è accettabile un maggior rumore di quantizzazione (grazie alle caratteristiche del nostro sistema psicovisivo) e quindi è possibile quantizzare più grossolanamente i coefficienti, riducendo di conseguenza il numero di bit da trasmettere. Infine i blocchi corrispondenti a contorni sono quelli più critici, danno origine a molti coefficienti non nulli ed il rumore di quantizzazione è percepibile.

Il decodificatore opera la trasformazione inversa, ricostruendo i campioni video (o le differenze) relative a ciascun blocco.

Il sistema è complesso, ma sfrutta in modo efficiente le ridondanze temporali, spaziali e statistiche del segnale video, minimizzando il degradamento percepito.

Bibliografia

1. N. Ahmed, T. Natarajan, K.R.Rao: "Discrete Cosine Transform" IEEE Trans. Computers, Vol. C-23, Jan. 1974, pp. 90-93

2. Speciale Italia '90, Elettronica e Telecomunicazioni, No. 3, 1990.

3. M. Barbero, S. Cucchi, M. Stroppiana: "A Bit-Rate Reduction System for HDTV Transmission", IEEE Trans. on Circuits and Systems for Video Technology, Vol. 1, No. 1, Marzo 1991, p. 4.

4. M. Barbero, M. Hofmann, N. D. Wells: "DCT Source Coding and Current Implementation of HDTV", EBU Tech. Review, No. 251, Spring 1992, p. 45-54

5. M. Barbero, M. Stroppiana: "Video Compression Techniques and Multilevel Approaches", SMPTE Journal, Vol. 103, No. 5, maggio 1994, p. 335.

6. ETS 300 174: "Network Aspects (NA); Digital coding of component television signals for contribution quality applications in the range 34 - 45 Mbit/s" (Nov. 1992).

7. Recommendation ITU-T J.81: "Transmission of component-coded digital television signals for contribution-quality applications at the third hierarchical level of ITU-T Recommendation G.702" (1993).

Che cosa è, come funziona: Uno standard pervasivo (MPEG-2 video)

Parte I

ing. Marzio Barbero e
ing. Natasha Shpuza

1. Premessa

Dal momento in cui è stato normalizzato, il sistema MPEG-2 ha avuto una rapida diffusione ed è attualmente alla base della maggior parte dei sistemi, a livello mondiale, di diffusione digitale dell'informazione video.

Questa prima parte descrive gli algoritmi alla base della codifica video. La seconda parte, pubblicata nel prossimo numero, approfondirà aspetti legati alle applicazioni e alla qualità dei sistemi video compressi. Prossime schede avranno come oggetto il sistema MPEG-2 nel suo insieme e la codifica dell'audio.

2. Cenni storici

MPEG[Nota 1] nasce nel gennaio 1988 come Experts Group dell'ISO/IEC JTC 1/SC 2/WG 8 e diventa SC 29/WG 11 nel novembre 1989.

All'inizio vennero definite tre fasi di lavoro, caratterizzate dal bit-rate complessivo necessario per la codifica di immagini in movimento per la memorizzazione digitale (DSM, *Digital Storage Media*) a 1-1,5 Mbit/s, a 1,5-5 Mbit/s e 5-60 Mbit/s.

La prima fase intendeva definire uno standard utilizzabile per la memorizzazione di informazioni audiovisive sul disco ottico allora disponibile il CD (il Compact Disc è caratterizzato da una capacità di 680 MB ed in quegli anni i lettori erano in grado di trasferire un flusso di dati appunto dell'ordine di 1,5 Mbit/s). Lo standard ISO/IEC 11172, noto come MPEG-1, ha quindi il titolo "Codifica di immagini in movimento e audio associato per media di memorizzazione digitale operanti fino a circa 1,5 Mbit/s", ma il riferimento al bit-rate è contenuto solamente nel titolo, infatti lo scopo era di definire uno "standard generico", in cui fosse specificato il minimo necessario, in pratica la sintassi e la struttura dei dati memorizzati o trasmessi, per consentire la decodifica dell'informazione audiovisiva, e lasciare ampia possibilità all'industria di ottimizzare i codificatori e i decodificatori.

La codifica video MPEG-1 è specificata nella parte 2 (11172-2:1993) ed basata sull'uso della DCT e della compensazione del movimento, analogamente allo standard ITU H.261 pubblicato nel 1990 per la videotelefonia e videoconferenza.

La seconda fase intendeva definire, malgrado l'obiettivo indicato nel titolo fosse limitato ai mezzi di memorizzazione digitale, una codifica adatta per le applicazioni diffusive (radiodiffusione e distribuzione via cavo) e dette origine allo standard ISO/IEC 13818, noto come MPEG-2. La prima sessione di MPEG-2 ebbe luogo nel luglio 1990, quando erano ancora in corso i lavori di definizione di MPEG-1. Nel 1992 fu eliminato il limite in termini di bit-rate, che intanto era passato da 5 Mbit/s a 10 Mbit/s, e la terza fase, inizialmente prevista per consentire la codifica di immagini in alta definizione (HDTV), confluì

Nota 1 - MPEG è l'acronimo di Motion Picture Expert Group, ovvero il WG 11 (Working Group 11, “Coding of Moving Pictures and Audio”) dello SC 29 (Subcommittee 29, “Coding of audio, picture, multimedia and hypermedia information”) del comitato congiunto JTC 1 (Joint Technical Committee 1, “Information Technology”) dell'ISO e IEC (International Organisation for Standardisation e International Electrotechnical Commission).
Il *Convenor* del gruppo è stato, fin dalla sua nascita Leonardo Chiariglione, dello CSELT (oggi TelecomItaliaLab), di Torino.

nella seconda fase (ciò spiega la mancanza di uno standard MPEG-3). Lo standard MPEG-2 video è del 1995 e coincide con lo standard ITU-T H.262.

Alla fine del 1996 MPEG-2 video fu scelto come base per il sistema televisivo digitale terrestre da introdurre in USA (ATV, *Advanced TV*). Fu scelto successivamente per la diffusione digitale da satellite, sempre in USA, per il progetto Direct TV. L'iniziativa europea per la diffusione digitale (DVB, *Digital Video Broadcasting*) diventò il maggior utilizzatore mondiale di MPEG. Seguì il DVD Forum e infine anche in Giappone MPEG-2 fu introdotto per la diffusione digitale sia di immagini in definizione standard (SDTV) che in alta definizione (HDTV).

Nel gennaio 1996 è stato definito il sistema di codifica per applicazioni professionali, in grado di operare sul video conforme alla Rac. ITU-R BT.601 senza apprezzabili degradamenti (*422 Profile*).

3. *Toolkit, Profile, Level*

Nella scheda precedente si è introdotto il termine *toolkit,* ovvero l'insieme degli utensili che vengono adottati per ridurre la ridondanza presente nell'informazione video e per eliminare eventuale informazione senza introdurre difetti rilevanti sull'immagine.

MPEG-2 adotta tutti i *tool* descritti: la trasformata coseno discreto (DCT) per sfruttare la correlazione spaziale, la codifica delle differenze fra quadri adiacenti e la compensazione del movimento per sfruttare la correlazione temporale, i codici a lunghezza variabile (VLC) per ridurre la ridondanza statistica, la matrice di visibilità e la quantizzazione dei coefficienti DCT per eliminare l'informazione meno rilevante.

Un sottoinsieme degli utensili, o tutti, concorrono alla definizione di un *profile*, un profilo specifico per una tipologia di applicazioni.

Nell'ambito di un singolo *profile,* possono essere definiti uno o più *level*, livelli di definizione dell'immagine, in termini di campioni e righe attive.

I concetti di *tool*, *profile* e *level* vengono adottati in MPEG-2 per consentire di comprendere in un unico standard generico i requisiti individuati durante il processo di normalizzazione.

In tabella 1 sono riassunte le caratteristiche principali dei profili e livelli previsti da MPEG-2. Quello più diffuso è il MP@ML, ovvero il profilo principale (*Main Profile*) al (@) livello principale (*Main Level*), è alla base dei sistemi DVB e DVD. Il 422P@ML è usato in produzione e postproduzione.

3.1 Main Profile

Il MP prevede più livelli, corrispondenti a dimensioni massime delle *picture,* si va dall'alta definizione, formato studio, al formato SIF, quello adottato in MPEG-1.

E' comunque il livello principale, cioè MP@ML quello più diffuso, in grado di codificare immagini basate sulla Rac. ITU-R BT.601, caratterizzate da un sottocampionamento della crominanza in verticale (convenzionalmente indicata come 4:2:0). Il decoder deve essere in grado di decodificare flussi con bit-rate fino ad un massimo di 15 Mbit/s.

3.2 4:2:2 Profile

E' stato definito per applicazioni professionali: ad esempio per la videoregistrazione in studio o per la trasmissione su rete di contribuzione o di distribuzione primaria. Opera quindi su un segnale conforme alla Rac. ITU BT.601 ed è caratterizzato da un bit-rate massimo (50 Mbit/s) che consente una qualità molto elevata (quasi la trasparenza) del segnale anche nel caso di più co-decodifiche in cascata.

3.1 Simple Profile

Il SP è praticamente identico al MP@ML tranne per il fatto di non prevedere l'uso di *B-picture*, la cui funzione è spiegata nel seguito, e di conseguenza limita l'uso di memorie di quadro video e riduce i costi realizzativi.

profili e livelli	SIMPLE 4:2:0	MAIN 4:2:0	422P 4:2:2 e 4:2:0	SNR 2 layer 4:2:0	SPATIAL 3 layer 4:2:0	HIGH 3 layer 4:2:2 e 4:2:0
HIGH 60 frame/s		1920x1088 80 Mbit/s I,P,B	1920x1088 300 Mbit/s I,P,B			1920x1152 100 Mbit/s I,P,B
HIGH-1440 60 frame/s		1440x1152 80 Mbit/s I,P,B			1440x1152 60 Mbit/s I,P,B	1440x1152 80 Mbit/s I,P,B
MAIN 30 frame/s	720x576 15 Mbit/s I,P	720x576 15 Mbit/s I,P,B	720x608 50 Mbit/s I,P,B	720x576 15 Mbit/s I,P,B		720x576 20 Mbit/s I,P,B
LOW 30 frame/s		352,288 4 Mbit/s I,P,B		352x288 4 Mbit/s I,P,B		

Tab. 1 - Tabella riassuntiva dei principali parametri che caratterizzano profili e livelli previsti dallo standard MPEG-2.

3.3 SNR Scalability Profile

I sistemi analogici sono caratterizzati da quello che in inglese è indicata come *graceful degradation*, ovvero la qualità dell'immagine e del servizio degradano progressivamente al peggioramento delle condizioni del canale. Nei sistemi digitali di trasmissione del video, invece, la qualità è praticamente costante, indipendente dalle condizioni del canale, fino a quando le tecniche di protezione e mitigazione degli errori non riescono più a far fronte alla crescita della probilità di errore e, in modo brusco, si passa in condizioni di mancanza del servizio.

Nel caso del profilo SNR si vuole ottenere un comportamento simile alla *graceful degradation* e i dati sono assegnati a due strati (*layer*) del flusso informativo. Il *layer* base contiene l'informazione indispensabile ad una decodifica completa dell'immagine, anche se quantizzata in modo più grossolano e quindi caratterizzata da un rapporto segnale/rumore inferiore; in genere è trasportato da un canale più protetto, caratterizzato da una QoS (*Quality of Service*) più elevata, garantendo quindi il livello minimo di qualità. L'informazione aggiuntiva del secondo *layer* consente, in condizioni ottimali del canale, di decodificare l'immagine alla massima qualità.

3.4 Spatial e Temporal Scalability

In un sistema scalabile spazialmente il flusso di dati è diviso in un *layer* base sufficiente a decodificare immagini a più bassa risoluzione (ad esempio SDTV, a definizione convenzionale) e in un ulteriore strato, che contiene l'informazione aggiuntiva che consente di ricostruire l'immagine originaria, a definizione più elevata (quella associata al livello HDTV-1440, cioè caratterizzata da una risoluzione doppia sia in orizzontale che in verticale).

Nel caso di scalabilità temporale un *layer* è relativo ad una ripetizione di immagine pari a 25 o 30 Hz, mentre un ulteriore *layer* contie-

ne le informazioni necessarie a riprodurre le immagini a 50 o 60 Hz.

3.5 MultiView Profile

Il MVP è stato approvato nel luglio 1996 per la codifica in modo efficiente di sequenze video ottenute da due telecamere che riprendono la stessa scena. In questo caso il *layer* base contiene una delle immagini ed è utilizzato per fornire la predizione ad un secondo *layer*, in modo da consentire la riproduzione di un video stereoscopico, su apposito ricevitore dotato di display stereroscopico.

4. Organizzazione dell'informazione video

4.1 Formati d'immagine, a scansione progressiva o interlacciata

Lo standard H.261, sviluppato per la videoconferenza su linee ISDN, era in grado di codificare immagini in formato CIF (*Common Interface Format,* 352 pixel x 288 righe a 30 quadri al secondo), un compromesso derivato dai due formati interlacciati previsti dalla Rac. ITU-R BT.601. Il compromesso, e la conseguente complicazione dei terminali, rese difficoltosa la diffusione del sistema.

MPEG-1 introdusse il quindi il formato SIF (*Standard Image Format*) che è ricavato sottocampionando sia orizzontalmente che verticalmente i due formati previsti dalla ITU-R BT.601 e ottenendo in un caso immagini progressive da 352 pixel per 288 righe a 25 quadri al secondo e nell'altro caso immagini da 352 pixel x 240 righe e 30 quadri al secondo. Il sottocampionamento verticale consiste semplicemente nella non codifica di uno dei semiquadri. In questo modo, se si opera nell'ambito di uno dei formati, telecamere e display sono di facile reperimento e il codec deve effettuare sottocampionamenti e interpolazioni non particolarmente complessi.

MPEG-2 MP@ML è in grado di codificare immagini a piena risoluzione (720 pixel x 576 righe per 25 quadri al secondo, oppure 720x480 per 30 quadri al secondo), sia nel caso di scansione progressiva che nel caso di scansione interlacciata. Nel caso di scansione interlacciata i due semiquadri (*field*) possono essere codificati contemporaneamente, come una singola immagine o *picture* (modalità *frame picture*) oppure possono essere codificati come due *picture distinte* (modalità *field picture*). Entrambe le modalità possono essere utilizzate all'interno della stessa sequenza.

Fig. 1 - Nella struttura di campionamento denominata 4:2:0 i campioni di crominanza C_R e C_B sono ottenuti mediante filtraggio e sottocampionamento a partire dal 4:2:2 e, a causa della struttura interlacciata, non sono in posizione equidistante rispetto ai campioni d'origine. Nella terminologia MPEG i due semiquadri video sono denominati *top field* e *bottom field*.

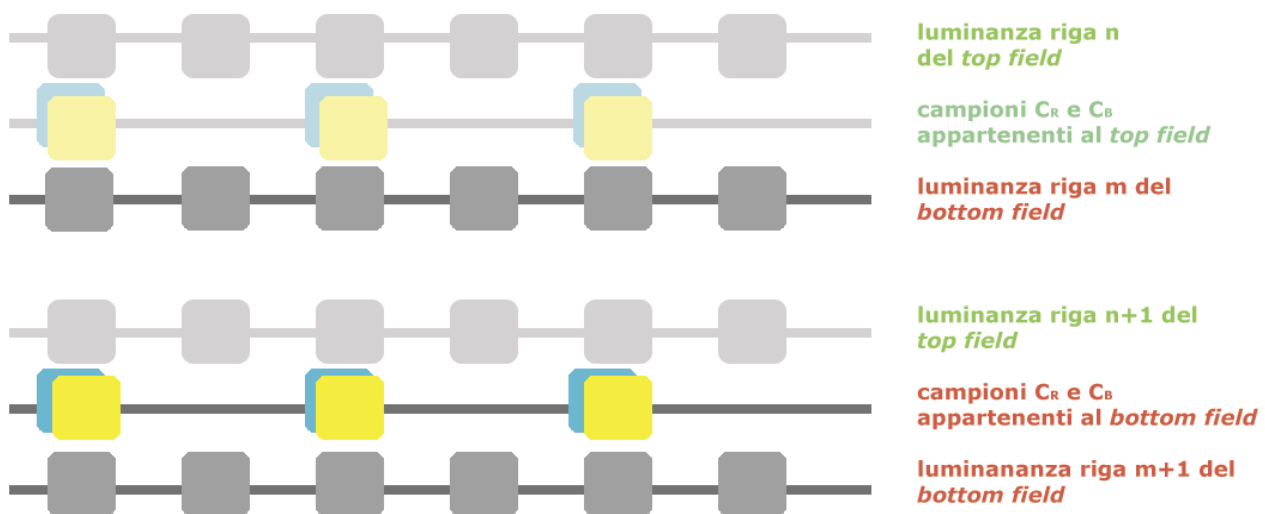
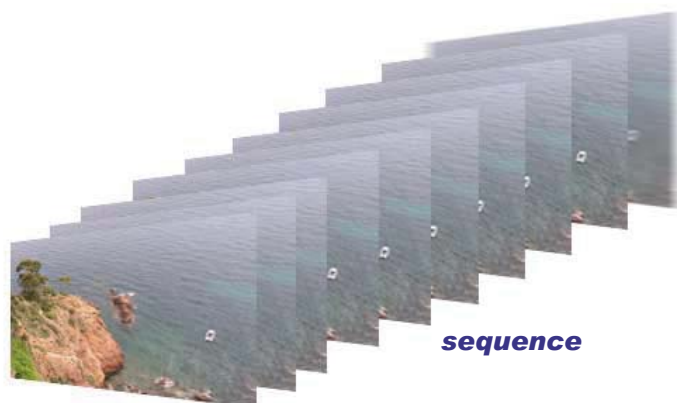
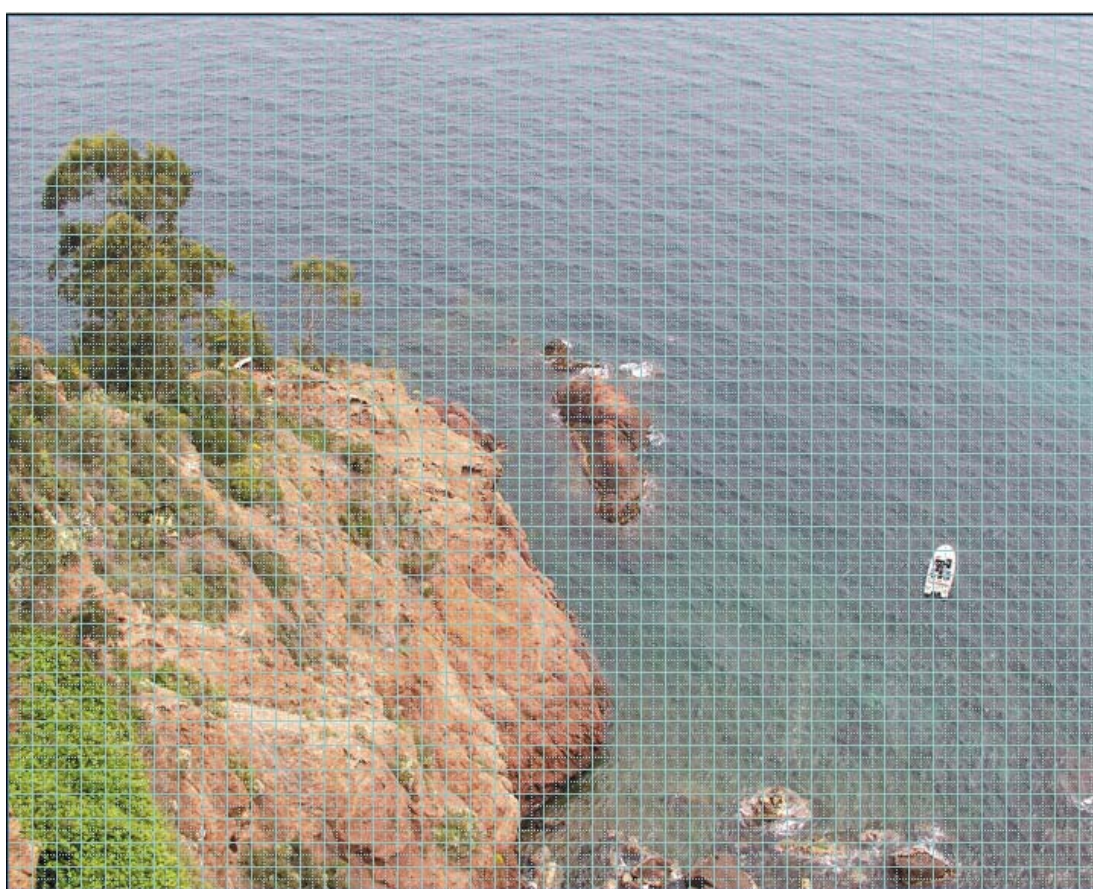
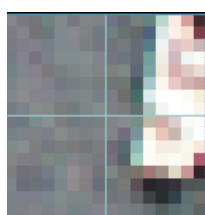
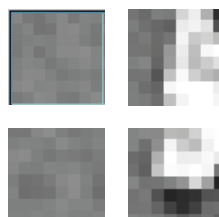
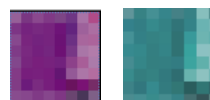



Fig. 2- Una *sequence* è una successione di *picture* video. Nel caso del profilo principale a livello principale MP@ML, ogni *picture* è costituita da un massimo di 720 campioni (*sample*) di luminanza per ciascuno delle 576 righe video (valore massimo). La *picture* è organizzata in fette (*slice*) costituiti da *macroblock* contigui. Il macroblocco corrisponde alla porzione di immagine a cui può venire applicata la compensazione del movimento ed è composto da 16x16 elementi di immagine. E' a sua volta formato da 4 blocchi (*block*) costituiti da 8x8 campioni (*sample*) di luminanza, mentre per quanto riguarda le due componenti di crominanza, queste sono sottocampionate sia in orizzontale che in verticale (struttura 4:2:0): a 16x16 campioni di luminanza corrispondono quindi due blocchi 8x8 di crominanza (uno per ciascuna delle due componenti C_R e C_B).

picture

slice

4.2 Campionamento 4:2:0

La Rac. ITU-R BT.601 prevede le due strutture di campionamento 4:4:4 e 4:2:2, e il 422 *Profile* dello standard MPEG-2 consente di operare direttamente sui campioni della struttura 4:2:2.

Per gli altri profili si prevede invece un ulteriore sottocampionamento delle due componenti di crominanza, riducendone la risoluzione verticale: questa struttura di campionamento viene convenzionalmente indicata come 4:2:0 (figura 1). In questo modo si ha una riduzione del 25% del numero di campioni su cui effettuare la compressione, ma si introduce un degradamento che è particolarmente critico nel caso di co-decodifiche successive, nel caso dei formati interlacciati. Tali profili sono quindi adatti ad applicazioni relative alla distribuzione all'utente finale.

4.3 Dalla *Sequence* al *Sample*

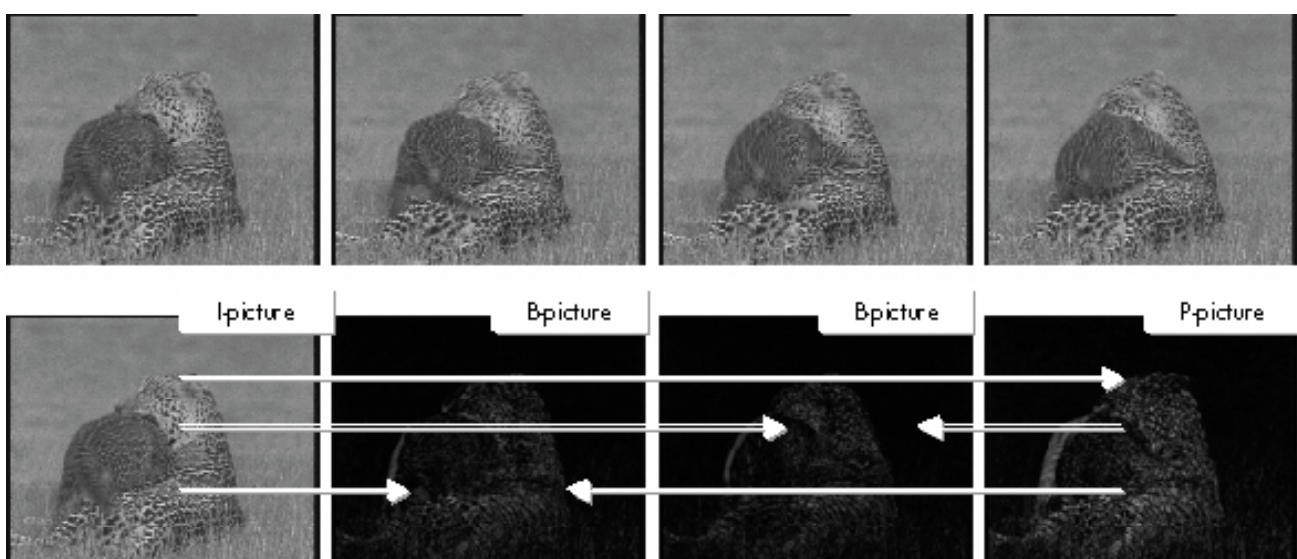
Il video in ingresso ad un codificatore MPEG è considerato una struttura organizzata secondo quanto indicato in figura 2.

La sequenza è una successione di *picture* video. Ciascuna *picture* è suddivisa in *slice*, fette costituite da un numero arbitrario di *macroblock* consecutivi, i macroblocchi sono raggruppamenti di 16 elementi di immagine in orizzontale per 16 elementi in verticale. Ogni macroblocco è a sua volta costituito da 4 *block*, blocchi di 8 per 8 *sample*, campioni di luminanza, più 2 blocchi 8x8 per ciascuna delle due componenti di crominanza (nei formati 4:2:0).

4.2 *Picture* di tipo I, P e B: il GOP

La DCT viene applicata ai singoli blocchi 8x8. Se il macroblocco è codificato in modo I (*Intra*), cioè sfruttando la sola ridondanza spaziale, la DCT viene applicata direttamente ai campioni, di luminanza o di crominanza, corrispondenti ai 6 blocchi che costituiscono il macroblocco. Se l'intera *picture* è codificata sfruttando esclusivamente la correlazione spaziale, viene denominata *I-picture* e tutti i macroblocchi sono di tipo I.

Se si sfrutta la ridondanza temporale, i coefficienti della DCT sono calcolati sulle differenze fra i campioni e i valori di predizione ottenuti a partire dai campioni di una *picture* precedente, eventualmente utilizzando anche la compensazione del movimento. Le immagini che contengono macroblocchi codificati sfruttando la predizione temporale sono denominate *P-picture* (*P=Predicted*). Una *P-picture* contiene normalmente anche dei blocchi di tipo intra, se non è stata individuata una buona predizione

Fig. 3 - In questo esempio quattro *picture* sono codificate, la prima come *I-picture,* la quarta come *P-picture* e le due intermedie come *B-picture*. Le differenze fra i valori dei campioni e le predizioni (le frecce indicano da quale *picture* sono ottenute le predizioni) sono inferiori nel caso delle B-picture, con un conseguente risparmio in termini di bit-rate.

temporale per tali porzioni d'immagine.

MPEG prevede un'ulteriore tipologia di immagini, le *B-picture,* per i macroblocchi delle quali i valori di predizione sono ottenuti a partire da *I-picture* o *P-picture* che precedono e seguono (*B=Bidirectionallly predicted*) quella attuale.

Poiché non risulta facile predire il futuro, per poter codificare i macroblocchi delle B-picture occorre avere delle memorie di quadro in modo che, con un opportuno ritardo, il codificatore abbia contemporaneamente a disposizione sia la *picture* di tipo I o P che precede che la *picture* di tipo I o P che segue (figura 3). Periodicamente è forzata la codifica di una *I-picture* in modo che il decodificatore al momento dell'accensione o a seguito di errori sul canale di trasmissione, possa recuperare la corretta decodifica della sequenza.

Una *I-Picture* è la prima di un gruppo di immagini denominato GOP (*Group of Pictures*). La lunghezza del GOP non è specificata da MPEG. Nel caso in cui non si utilizzi la correlazione temporale, le *picture* sono tutte di tipo I e la lunghezza del GOP è pari a 1. Si possono costruire strutture di GOP composte solo dalla *I-picture* seguita da una o più *P-picture*. Se vi sono *B-picture*, la struttura tipica è quella denominata IBBP, cioè vi è una coppia di *B-picture* preceduta e seguita da *I-* o *P-picture*.

In genere una *I-Picture* richiede, per la codifica, una quantità di dati superiore rispetto a quelle predette, per cui si ha un risparmio in termini di bit-rate per GOP lunghi, comunque tipicamente la lunghezza del GOP non supera 12 o 15, ma questo valore non è necessariamente fisso e non è indicato da nessun parametro all'interno del flusso binario: la scelta è operata dal codificatore e il decodificatore si adatta di conseguenza.

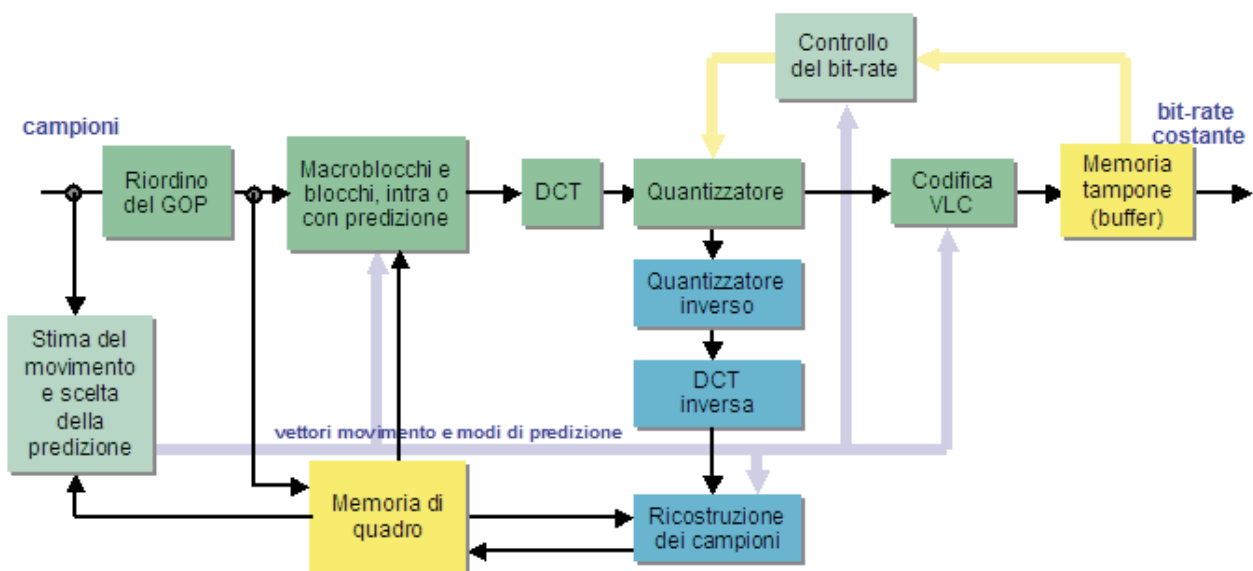
5. La codifica

5.1 Schema del codificatore

Lo schema di un codificatore MPEG è riportato in figura 4. Il processo di codifica segue lo schema indicato in figura 5.

Il video viene organizzato in funzione della struttura a GOP prescelta.

Nel caso in cui la *picture* è di tipo P o B, viene determinato il macroblocco di predizione ottimale, sfruttando la compensazione del

Fig. 4 - Schema a blocchi di un codificatore MPEG.

movimento.

A tutti i blocchi 8x8 (di luminanza e crominanza) viene applicata la DCT. Nel caso dei macroblocchi di tipo I la precisione dei coefficienti è fino a 10 bit nel caso del MP@ML, mentre nel caso del 422P arriva a 11 bit.

Nel caso in cui i macroblocchi siano di tipo P o B, la DCT è applicata sulle differenze fra i campioni video attuali e quelli predetti a partire dai campioni contenuti nelle memorie di quadro. Nelle memorie di quadro sono memorizzati i valori delle *picture* ricostruite localmente, ma identiche (a parte gli eventuali errori di trasmissione) a quelle presenti nei decodificatori.

Ai coefficienti appartenenti ai blocchi sono applicate le matrici di pesatura (figura 6).

La quantizzazione dei coefficienti DCT è la principale causa di perdita di informazione (e di qualità dell'immagine) nella codifica MPEG. Per ogni macroblocco viene trasmesso il parametro *mquant* che indica il fattore di scala con cui vengono codificati i coefficienti, il codificatore opera la scelta per ottenere la massima qualità (il minimo errore di quantizzazione) con il bit-rate disponibile.

Successivamente si effettua la scansione di tipo zig-zag (figura 7) in modo da operare sulla sequenza dei 64 coefficienti generati per ciascun blocco.

La parole a lunghezza variabile vengono assegnate in base a due informazioni: il numero di zeri che precede un coefficiente non nullo (*run*) e il valore di tale coefficiente (*level*). In base alla coppia *run+level* viene generata una parola VLC la cui lunghezza può variare da 1 a 16 bit + segno. Molte combinazioni *run+level* non sono associate ad una parola VLC, ma sono codificate come *run* (6 bit) + *level* (12 bit). Una parola specifica (EOB, *end-of-block*) indica che tutti i coefficienti del blocco successivi a quello precedentemente codificato sono nulli.

5.2 Predizioni

Il predittore normalmente basa la stima del

Stima del movimento

Trasformata Coseno Discreta (DCT)

Matrice di pesatura e Quantizzazione

Scansione a zig-zag

Codifica con parole a lunghezza variabile (VLC) delle sequenze di zeri e dei livelli

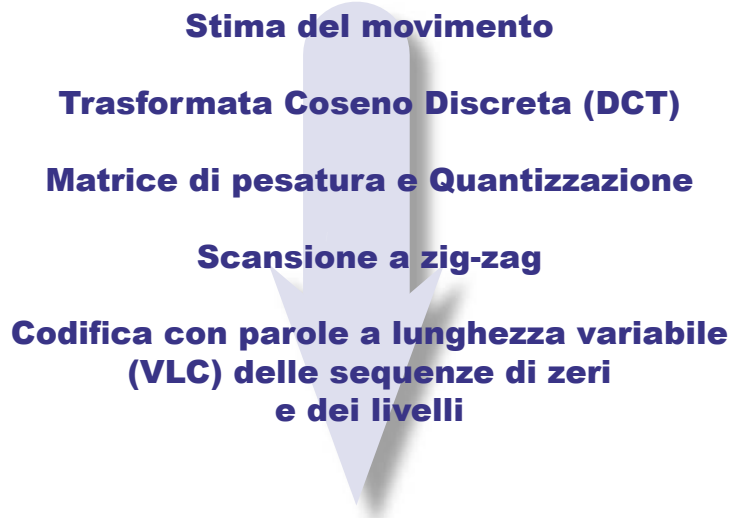



Fig. 5 - Il processo di compressione è basato, nell'ordine, in: stima del movimento (applicata solo nel caso delle *picture* B e P), trasformata DCT applicata ai blocchi 8x8, quantizzazione, scansione dei coefficienti e codifica VLC.

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

Fig. 6 - La matrice di pesatura di sinistra è utilizzata per dividere i 64 coefficienti nel caso di blocchi di tipo I, mentre la matrice di destra è utilizzata nel caso dei blocchi di tipo P e B. Sono applicate sia ai blocchi di luminanza che a quelli di crominanza.

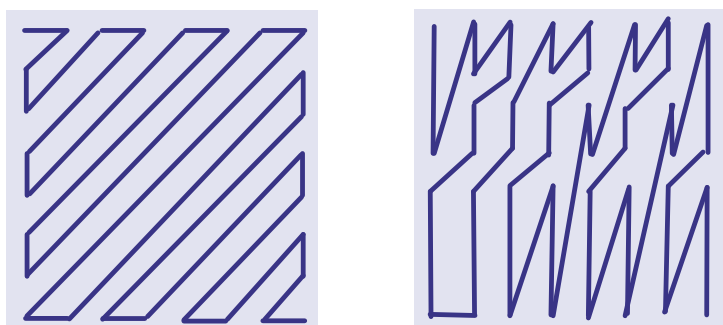
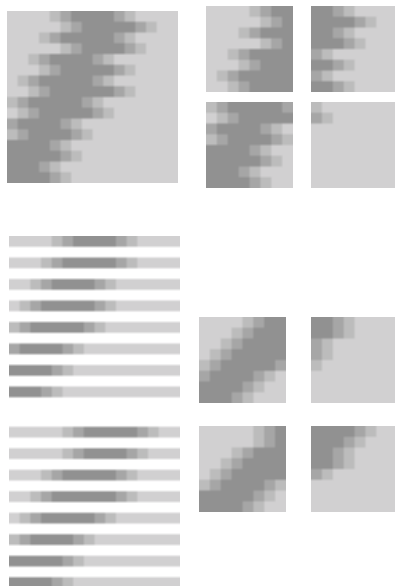



Fig. 7 - I 64 coefficienti DCT sono trasmessi secondo un percorso a zig-zag, dapprima quello relativo alla componente continua e via via quelli relativi alle frequenze spaziali più alte. Il percorso di sinistra è quello utilizzato nel caso di codifica di video in formato progressivo, quello di destra è invece utilizzato nel caso di video in formato interlacciato.

Fig. 8 - Nel caso di video interlacciato, è possibile scegliere se codificare separatamente la luminanza relativa ai due semiquadri.
Nel caso di movimenti veloci è possibile che i blocchi ottenuti codificando separatamente le otto righe video appartenenti al *top field* e le otto righe video appartenenti al *bottom field* presentino una correlazione maggiore rispetto alla codifica nei quattro blocchi ottenuti combinando i due *field* in un *frame*. Nell'esempio sono messe a confronto le due possibilità: il *frame* origina quattro blocchi (in alto) che presentano una minore correlazione rispetto ai quattro blocchi originati dai due *field* considerati separatamente (in basso).

movimento sulle immagini memorizzate ottenute grazie alla decodifica locale e la stima è fatta con una precisione pari a mezzo pixel: la predizione ottimale sfrutta al meglio la ridondanza temporale e riduce l'effetto del rumore presente sull'immagine, grazie alle operazioni di media (e quindi filtraggio) su più pixel.

Sono previsti differenti tipi di predizione.

Nella *field prediction*, le predizioni sono ottenute indipendentemente per ciascun semiquadro usando i dati memorizzati relativi ad uno o più semiquadri decodificati. Nelle *field picture* tutte le predizioni sono di tipo *field*.

Nella *frame prediction*, le predizioni sono ottenute da uno o più quadri precedentemente decodificati e la selezione *field prediction* o *frame prediction* è effettuata a livello del singolo macroblocco.

Nel caso di movimenti veloci, ci possono essere vantaggi nel codificare separatamente la luminanza relativa ai due semiquadri (figura 8): per ciascun semiquadro sono originati due blocchi di luminanza ed un vettore movimento (per questi macroblocchi vengono dunque trasmessi due vettori movimento, ciascuno relativo a 16x8 campioni).

I modi di predizione prescelti vengono codificati a lunghezza variabile e inviati al decoder.

Analogamente vengono codificati con VLC i vettori movimento: nel caso di MP@ML i valori codificabili sono da -1024 a +1023,5 in orizzontale e da -128 a +127,5 in verticale.

5.3 Qualità costante o bit-rate costante (VBR, CBR)

Si è visto che la perdita in qualità è legata soprattutto all'uso di un fattore di quantizzazione *mquant* più elevato. Al suo crescere diminuisce la precisione con cui vengono trasmessi i coefficienti DCT e conseguentemente aumenta il rumore di quantizzazione, ma si riduce il bit-rate: infatti in ogni blocco vi sono più coefficienti nulli ed il livello di quelli non nulli si riduce, consentendo l'utilizzo di parole VLC più brevi.

Pertanto se si mantiene costante il valore di *mquant*, le immagini sono codificate a qualità costante, ma il bit-rate varia in funzione della criticità dell'immagine rispetto alla codifica.

Per certe applicazioni è conveniente utilizzare una codifica a qualità costante, e conseguente a bit-rate variabile: la codifica VBR (*variable bit-rate*) è spesso adottata per il video memorizzato, ad esempio su supporto ottico (DVD).

In altri casi invece è vincolata la velocità, ad esempio perchè l'informazione deve essere trasferita su un canale a bit-rate costante (CBR, *costant bit-rate*).

In questo caso diventa determinante il ruolo della memoria tampone (*buffer*) indicato nello schema di figura 4. il bit-rate medio dei dati in ingresso deve coincidere con quello all'uscita del *buffer,* costante ed imposto dal canale. Per ottenere tale risultato *mquant* viene variato dinamicamente in funzione del livello di riempimento del buffer. Se il buffer tende a riempirsi, i coefficienti DCT vengono quantizzati più grossolanamente, il flusso di dati in ingresso diminuisce e il buffer tende a svuotarsi. Quando il livello di riempimento è basso, si migliora la quantizzazione e il livello torna a salire.

L'ottimizzazione delll'algoritmo di gestione del buffer deve tenere in considerazione il fatto che le I-picture generano una quantità di dati molto superiori a quelle predette. Le B-picture invece possono contribuire molto poco alla quantità di dati complessiva, anche perché spesso si accetta per esse una quantizzazione più grossolana, poiché il sistema psicovisivo umano è considerato poco sensibile alle fluttuazioni di qualità video se avviene con frequenza prossima alla periodicità IBBP.

6. Video elementary stream

Il risultato della codifica video è un flusso di dati binari denominato *Elementary Stream* (ES): esso contiene tutta l'informazione necessaria a decodificare un segnale video.

La figura 9 è una rappresentazione schematica dell'organizzazione dei dati presenti nello *stream* e può essere utilizzata per ricapitolare brevemente gli algoritmi e le funzioni precedentemente descritti.

A livello superiore troviamo la *sequence,* l'intestazione (*sequence header*) contiene le informazioni di base, necessarie per iniziare la decodifica, quali le dimensioni delle *picture,* il formato dell'immagine (*aspect ratio*), la frequenza di quadro, le tabelle di quantizzazione. L'intestazione, data la sua importanza, è ripetuta periodicamente (ad esempio due volte al secondo).

Raggruppati nella sequenza vi sono i *Group Of Pictures*. L'intestazione (*GOP header*) contiene le informazioni necessarie per la riproduzione temporalmente corretta del video (*time code*) e alcuni flag utilizzati nell'editing.

Le *picture* costituiscono ciascun GOP. L'intestazione (*picture header*) contiene un riferimento temporale, l'indicazione del tipo di *picture* (I,P,B).

La *picture* è costituita da *slice*. L'intestazione (*slice header*) è identificata, come le altre intestazioni, da un codice (*start code*) che non può essere duplicato all'interno del flusso. E' l'entità minima, all'interno del flusso elementare video, grazie alla quale è possibile ottenere la sincronizzazione e quindi la corretta decodifica. In genere una *slice* corrisponde ad un insieme di macroblocchi pari a 16 righe video, ma in applicazioni in cui occorra una veloce e sicura sincronizzazione, è possibile avere più *slice*, al limite una *slice* in corrispondenza di ciascun macroblocco.

L'intestazione del macroblocco (*macroblock header*) contiene tutta l'informazione necessaria a decodificare correttamente la porzione di immagine (64 elementi di immagine) che lo costituiscono: l'indirizzo spaziale all'interno dell'immagine, i vettori movimento, i modi di predizione e di trasformazione (*field/frame*), il fattore di quantizzazione.

Seguono i coefficienti DCT codificati VLC (*run+level*) e le parole EOB (*End Of Block*) separano i quattro blocchi (*block*) di luminanza e i due di crominanza che (nel MP@ML) costituiscono un macroblocco.

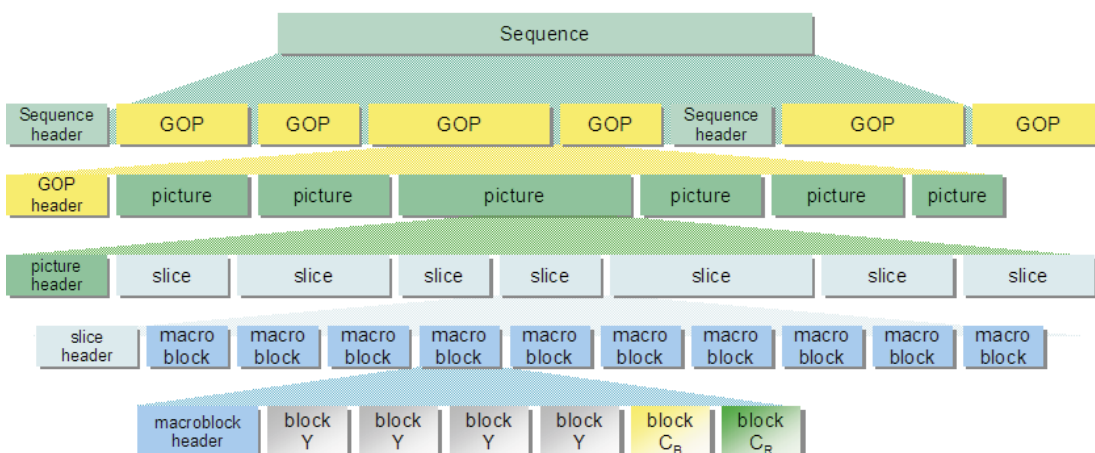



Fig. 9 - Schema dell'organizzazione del *video elementary stream*.

Che cosa è, come funziona: Advanced Video Coding (AVC - H.264): Il prossimo futuro

Parte I

ing. Marzio Barbero e
ing. Natasha Shpuza

1. Premessa

La definizione della codifica video MPEG-2 risale alla prima metà degli anni novanta, negli ultimi anni gli esperti hanno proseguito il loro lavoro per ottenere codifiche sempre più efficienti, soprattutto per le applicazioni a basso bit-rate. L'evoluzione degli algoritmi, ma soprattutto delle capacità di elaborazione integrabili nei *microchip* ha consentito la definizione del sistema AVC (*Advanced Video Coding*) che ha buone possibilità di trovare ampia applicazione nei prossimi anni.

Lo standard specifica la codifica del video (VCL, *Video Coding Layer*), che è oggetto di questa scheda, e il formato con cui organizzare i dati video per il trasporto e la memorizzazione (NAL, *Network Abstraction Layer*).

Fig. 1 - Il Joint Video Team comprende esperti di MPEG e dell'ITU-T e ha completato le specifiche che sono pubblicate separatamente dai due organismi di standardizzazione. Gli standard sono tecnicamente equivalenti, anche se non necessariamente identici.

2. Cenni storici

La codifica delle informazioni video è oggetto di studio dei gruppi di normalizzazione ISO/IEC (MPEG, *Motion Picture Expert Group*) e ITU (VCEG, *Video Coding Experts Group*), il cui lavoro portò alla definizione della parte 2 di MPEG-2 e allo standard ITU-T H.262 nel 1995. L'ITU sviluppò indipendentemente l'H.263 e due estensioni (pubblicate sotto forma di annessi) e denominate H.263+ e H.263++, mentre in MPEG si procedeva allo sviluppo della parte 2, relativa alla codifica video, dello standard MPEG-4 partendo, come base da H.263.

Nel 2001 fu deciso, per evitare divergenze nello sviluppo ed i problemi di sincronizzazione fra i due organismi di standardizzazione, di stabilire un gruppo congiunto, il JVT (*Joint Video Team*) per portare a termine il lavoro di definizione di un unico sistema di codifica video (figura 1).

Nella riunione MPEG-4 del marzo 2003 a Pattaya venne approvato il nuovo sistema di codifica, AVC (*Advanced Video Coding*), come parte 10 dello standard MPEG-4 ISO/IEC 14496-10. In ambito ITU lo standard, inizialmente indicato provvisoriamente come H.26L, sarà pubblicato come ITU-T H.264.

3. Algoritmi e profili

Lo standard AVC, così come avviene nel caso di MPEG-1 e MPEG-2, non definisce un CODEC (cioè codificatore e decodificatore, *COder* ed *DECoder*), bensì la sintassi del flusso dati (*stream syntax*) e il metodo di decodificarlo. I *tool*, cioè gli algoritmi, adottati, non sono sostanzialmente diversi da quelli illustrati per MPEG-2 video: la maggiore efficienza di codifica è dovuta alla cura dei dettagli di ciascun elemento funzionale (figura 2).

Lo standard supporta la codifica del video nel formato 4:2:0, interlacciato o progressivo. Una *sequence* è organizzata in *frame*, costituiti da due *field* (*top* e *bottom field*). Una sequenza progressiva è codificata come quadri (*frame picture*), consistenti in due semiquadri (*field*), acquisiti nello stesso istante di tempo.

Sono previsti differenti profili, indirizzati ad applicazioni differenti:

- *Baseline Profile*, destinato ad applicazioni in cui si richieda un ridotto ritardo dovuto alla codecodifica, ad esempio videotelefonia o applicazioni mobili, ...
- *eXtended Profile*, per applicazioni mobili e *streaming, ...*
- *Main Profile,* per applicazioni diffusive, formato video interlacciato, ...

4. La codifica

Per ciascun elemento funzionale nel seguito si descrivono brevemente i miglioramenti apportati in AVC rispetto ad MPEG-2, che possono essere sintetizzati in:

- applicazione della trasformata su blocchi più piccoli
- miglioramenti relativi alla valutazione e alla compensazione del movimento
- filtro di ricostruzione nel *loop* di decodifica per ridurre l'effetto di blocchettizzazione
- miglioramento della codifica entropica.

Predizione intra e inter con stima del movimento strutturata ad albero e precisione a 1/4 pel.
Filtro di ricostruzione nel loop di decodifica.

Trasformata applicata a blocchi 4x4

Matrice di pesatura e Quantizzazione

Riordino dei coefficienti

Codifica con parole a lunghezza variabile basata sul contesto (CAVLC o CABAC)

Fig. 2 - Il processo di compressione è analogo a quello applicato negli standard precedenti, in particolare MPEG-2, ma, a spese di una maggiore complessità, consente un migliore sfruttamento delle ridondanze statistiche e riduce la percezione soggettiva delle distorsioni.

4.1 Macroblock e slice

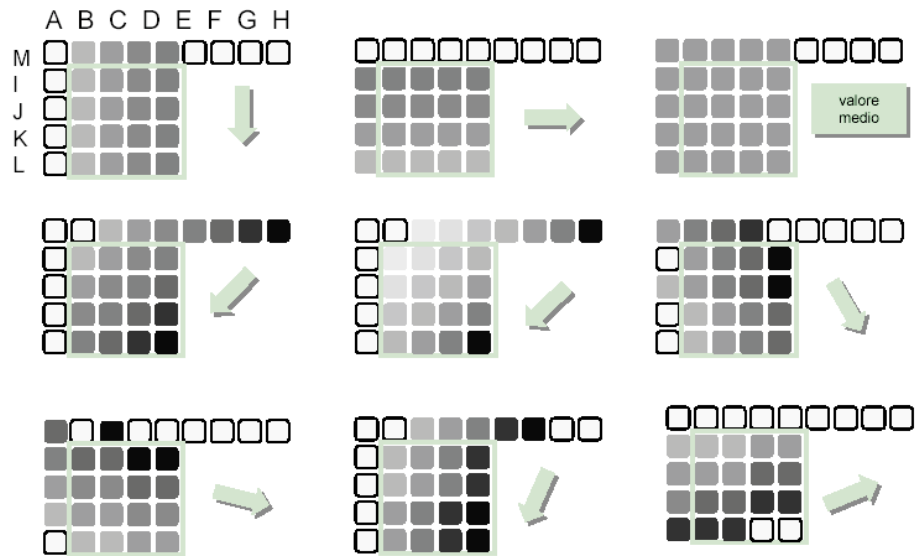
I macroblocchi sono anche in AVC costituiti da 16x16 elementi di immagine: 16x16 campioni (*sample*) di luminanza e 8x8 campioni per ciascuna componente di crominanza C_B e C_R.

I blocchi (*block*) sono costituiti da 4x4 campioni (un quarto della dimensione adottata in MPEG-2).

I macroblocchi sono organizzati in *slice*, un sottoinsieme di immagine decodificabile indipendentemente dalle altre. L'ordine di trasmissione dei macroblocchi non è necessariamente quello originario nell'immagine, ma è indicato dal codificatore in una apposita mappa (*Macroblock Allocation Map*).

Sono definiti 5 differenti tipi di *slice*.

I primi tre, analogamente a quanto visto per MPEG-2, sono I (*intra*), P (*predictive*) e B (*bi-predictive*) e le predizioni sono ottenute a parti-

Fig. 3 - Nella codifica intra le predizioni per i blocco 4x4 di luminanza sono determinate in base ai valori dei campioni qui indicati con le lettere A-M, precedentemente decodificati. Le frecce indicano la direzione di maggior correlazione e sulla cui base si determinano i valori di predizione.
Ad esempio, per il modo 0 (verticale) in alto a sinistra, i valori predetti sono ottenuti a partire da quelli precedentemente decodificati ABCD, mentre nel modo 2 (DC) in alto a destra, tutti i 16 valori sono ottenuti come media di quelli adiacenti al blocco, cioè A-D e I-L.

re dalle *picture* precedentemente codificate. In AVC più *picture* possono essere utilizzate per le predizioni e pertanto codificatore e decodificatore memorizzano le *picture* utilizzate per le predizioni in una apposita memoria (*multi-picture buffer*) e il controllo per la gestione del *buffer* è specificato nel flusso dati.

Nelle applicazioni di *streaming* via internet spesso lo stesso video è codificato a differenti bit-rate ed il decoder tenta di accedere al flusso a più elevato bit-rate, che fornisce una più elevata qualità, ma se le condizioni del canale non lo permettono, commuta al flusso a bit-rate più basso. Quando si utilizza MPEG-2 queste operazioni di commutazione possono essere effettuate a livello di GOP, in corrispondenza di una *I-picture* e ciò implica l'uso di GOP relativamente corti e lo sfruttamento non ottimale della ridondanza temporale dell'informazione video. In AVC sono stati pertanto definiti ulteriori due tipi di *slice*, denominati SI (*Switching I*) e SP (*Switching P*) che consentono un'efficiente commutazione fra flussi di dati a bit-

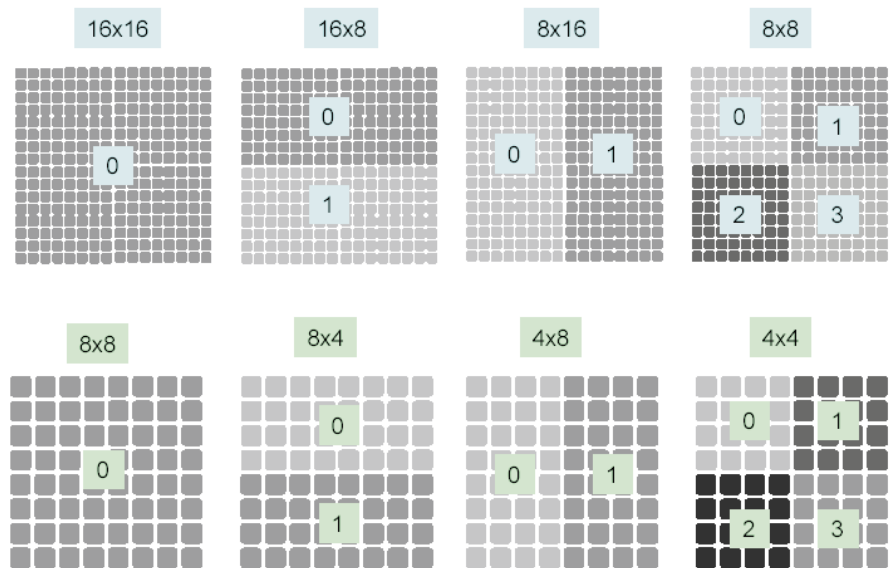



Fig. 4 - La segmentazione dei macroblocchi (luminanza) al fine della compensazione del movimento prevede i 4 modi indicati nella parte alta della figura.
Se è scelto il modo 8x8, ciascuna delle 4 partizioni possono essere ulteriormente suddivise in sottopartizioni, come indicato in basso nella figura.

rate differente, senza rinunciare al massimo sfruttamento della ridondanza temporale.

4.2 Predizione e codifica Intra

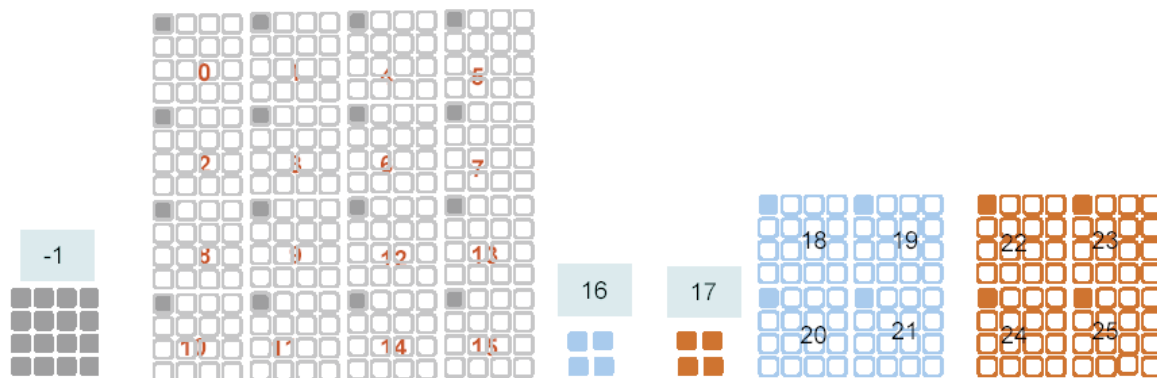
Nella codifica intra è sfruttata la sola correlazione spaziale: per aumentare l'efficienza vengono codificate le differenze fra i campioni del macroblocco e i campioni precedentemente codificati, tipicamente quelli posizionati sopra e a sinistra (figura 3) e sono definiti 9 modi distinti di predizione. Nel caso di aree piatte, con scarso dettaglio si può adottare la codifica intra sull'intera area 16x16 ed in tal caso sono definiti altri 4 modi di predizione per l'intero macroblocco.

4.3 Predizione e codifica inter

Nella codifica di tipo inter si parte da una predizione ottenuta, sfruttando la correlazione temporale, da uno o due quadri precedentemente codificati. La predizione può essere ottenuta mediante una stima ed una compensazione del movimento (*motion compesated prediction*). A differenza dagli standard precedenti, la dimensione del blocco su cui si effettua la predizione può variare da 16x16 fino a 4x4 (figura 4). Questo metodo di partizionare i *macroblock* in *sub-block* è denominato *tree structured motion compensation* e in fase di codifica sono possibili molteplici scelte che hanno implicazioni differenti sul numero di bit necessario a codificare i vettori movimento e le differenze residue: in genere dimensioni elevate del blocco sono convenienti in aree piatte, mentre in aree ricche di dettagli si può trarre vantaggio dall'uso di aree ridotte.

La precisione per i vettori movimento si incrementa da 1/2 elemento di immagine, utilizzato in MPEG-2, a 1/4 di elemento di immagine. Per ottenere questa precisione si utilizza un filtro digitale (6-*tap* FIR, *Finite Impulse Response*) che fornisce, a partire dalla somma pesata dei valori dei 6 campioni di luminanza adiacenti, i valori interpolati a 1/2 e una successiva interpolazione bilineare permette di ricavare i valori a 1/4. Nel caso della crominanza e di formato 4:2:0 la precisione è portata a 1/8, che corrisponde al valore 1/4 per la luminanza.

Esiste una correlazione fra i vettori movimento delle sottopartizioni adiacenti: essa è sfruttata calcolando un valore di predizione MVp dei vettori movimento relativi ad un macroblocco. Il valore di MVp è calcolato sia in codifica che in decodifica sulla base della struttura in termini di sottopartizioni che costituiscono il *macroblock*. In questo modo al decoder vengono inviati solo gli MVD, i valori delle

Fig. 5 - Ordine di trasmissine dei dati di un macroblocco. Se codificato nel modo 16x16 intra, tutti i coefficienti DC sono raggruppati nel blocco -1 trasmesso per primo. Vengono trasmessi i blocchi 0-15 (nel caso di modo intra i valori corrispondenti alla posizione DC sono posti a 0). Analogamente i blocchi 16 e17 contengono i coefficienti DC delle componenti di crominanza C_B e C_R, seguono i blocchi 18-25.

differenze fra i vettori movimento e il valore predetto MVp.

4.4 Trasformata e quantizzazione

Si utilizzano tre trasformate che dipendono dal tipo di dati che devono essere elaborati (figura 5):

- una trasformata 4x4 dei 16 coefficienti DC nel caso di macroblocchi intra 16x16
- una trasformata 2x2 per i coefficienti DC delle crominanze di tutti i macroblocchi
- una trasformata 4x4 di tutti gli altri dati differenze.

Il tipo di trasformata adottato è basato sulla DCT (*Discrete Cosine Transform*), ma sono state apportate delle modifiche affinché le operazioni richiedano somme e scalamenti effettuabili con numeri interi a 16 bit in modo da non avere perdita di precisione effettuando la trasformazione diretta seguita da quella inversa.

Esistono 52 passi di quantizzazione, denominati Q_p e questa ampia gamma di valori permette al codificatore di raggiungere il miglior compromesso fra qualità e bit-rate.

4.5 Filtro di ricostruzione

L'effetto di blocchettizzazione è uno dei degradamenti caratteristici delle tecniche di compressione che operano su macroblocchi di campioni video: è particolarmente visibile e fastidioso.

AVC introduce un filtro apposito che è applicato prima della trasformata inversa sia nel codificatore, prima della ricostruzione delle immagini utilizzate per le predizioni, sia nel decodificatore. Si ottengono due principali vantaggi: una minore visibilità dei bordi dei blocchi e una migliore predizione inter con compensazione del movimento (nel caso di predizione intra i macroblocchi sono filtrati, ma la predizione è ottenuta dai macroblocchi ricostruiti non filtrati).

4.6 Codifica VLC

I simboli che rappresentano i parametri relativi ai modi di codifica e predizione, i vettori movimento e i coefficienti della trasformata vengono codificati con codici a lunghezza variabile.

Lo standard specifica diversi tipi di codifica entropica: una codifica a lunghezza variabile (VLC, *Variable Lenght Coding*) basata su tabelle di assegnazione statiche oppure basate sul contesto CAVLC (*Context Adaptive Variable Lenght Coding*) e CABAC (*Context Adaptive Binary Arithmetic Coding*).

Il CAVLC utilizza diverse tabelle VLC specificatamente ottimizzate per i vari elementi sintattici in base a quelli precedentemente trasmessi. A seguito della predizione, trasformazione e quantizzazione i valori relativi ai coefficienti sono molto spesso nulli o molto piccoli: la codifica a lunghezza variabile sfrutta le sequenze di zero (codifica *run-level*), l'elevata frequenza di valori +1 e -1, e la correlazione fra il numero di coefficienti non nulli di un blocco e quello nei blocchi adiacenti.

Il CABAC, che è utilizzato nel *Main Profile*, sfrutta in modo ancora più efficiente la correlazione fra simboli perché utilizza la statistica dei simboli precedentemente codificati per stimare la probabilità condizionata, usata per selezionare uno fra i diversi modelli.